

2016 / 2017

# GUIDE DU BIG DATA

L'ANNUAIRE DE RÉFÉRENCE  
À DESTINATION DES UTILISATEURS

BIG DATA



# MATLAB PARLE LE MACHINE LEARNING

Avec MATLAB, créez des modèles prédictifs à l'aide de régressions, de clustering, de classification ou encore de deep learning et déployez-les dans votre environnement de production.

[mathworks.fr/machinelearning](http://mathworks.fr/machinelearning)

# DU BIG DATA A L'INTELLIGENCE DE LA DONNÉE

Cette année, vous l'aurez compris au travers de nombreuses chroniques, l'heure de la monétisation du Big Data a sonné ! Avec un marché du Big Data qui atteindrait 652 M€ à l'horizon 2018 en France (étude ATOS/IDC) et plus de 6,9 milliards en Europe de l'Ouest (étude IDC), la disponibilité des données n'est plus une problématique majeure.

L'enjeu réside dans la capacité des métiers et des spécialistes de la data à exploiter quotidiennement et de manière intelligente les datas pour tirer leur épingle du jeu.

Responsables de politiques publiques, Data Heroes au sein de leurs entreprises, Startups disruptives... cette année le Guide a donc souhaité vous présenter au travers d'interviews de fond les hommes et femmes qui mènent la transformation data.

Vous découvrirez leurs parcours, leurs positionnements au sein de l'organisation, leurs équipes, l'historique de mise en place de leurs projets et en quoi ils se positionnent en pionniers de l'Innovation Data.

Et, car l'inspiration n'est pas limitée aux frontières nationales, découvrez à travers ces témoignages d'experts français comme internationaux, les évolutions majeures qui marqueront le Big Data en 2017 :

DATA-DRIVEN **ECONOMY**  
MIXED DATA & SMART DATA  
BIG DATA AS **A SERVICE**  
PLATEFORME DATA, CLOUD et DIGITAL  
ECOSYSTEME DES **APIs**  
PROTECTION, ETHIQUE ET SÉCURITÉ DES DONNÉES  
DATA **ÉMOTIONNELLES !**

Nous avons enfin souhaité dédier un dossier spécial à l'Intelligence Artificielle, prolongement indéniable de l'interprétation intelligente des Big Data. En effet, **Prédictif, Machine Learning, Natural Language Processing, Bots...** sont sur toutes les lèvres. Découvrez donc le panorama dédié à l'évolution de l'Intelligence Artificielle et deux applications innovantes plus que remarquées dans le domaine !

Nous vous laissons donc picorer, sélectionner ou dévorer les interviews de cette 4ème édition du Guide selon vos problématiques et aspirations.

Bonne lecture !



**Marion PEAUDECERF**

Directrice de publication

+33 (0)1 84 83 03 03 / mpeaudecerf@corp-agency.com

Le Guide du Big Data est réalisé par la société Corp.

# LES TRIBUNES DU BIG DATA

## EDITO

03



**CHRISTIAN REIMSBACH-KOUNATZE**  
Les Enjeux de la Data-Driven Economy

06



**MARTA NAGY-ROTHENGASS**  
The framework conditions for the European Data Economy

11



**ALAIN BENSOUSSAN**  
Chronique juridique du Big Data, des algorithmes et de l'Intelligence Artificielle

15



**JÉRÔME BÉRANGER**  
De l'éthique dans l'exploitation et le traitement des données numériques

19



**BENOÎT BINACHON**  
Du recrutement dans la Data à la Data dans les processus de recrutement

24



**CHRISTOPHE MONTAGNON**  
RH & BIG DATA : Des outils au service de la fluidification du marché de l'emploi

28



**MARK VAN RIJMENAM**  
Mixed Data, BDaaS, APIs... Future trends in Big Data

32



**OLIVIER EZRATTY**  
Les avancées de l'Intelligence Artificielle

35



**FRANCISCO DE SOUSA WEBBER**  
Computable Language and Natural Language Processing

38



**CAROLINE CHOPINAUD / CLODÉRIC MARS**  
craft ai : L'Intelligence Artificielle as-a-service

41



**BÉATRICE TOURVIEILLE**  
Voyages-sncf.com : Digital et Big Data au service de la personnalisation

44

## ANNUAIRE ENTREPRISES

AB INITIO SOFTWARE .....	76	CLOUDERA FRANCE .....	104
ACCENTURE .....	78	CLUB DECISION DSI .....	106
ADVANCED SCHEMA .....	80	COHERIS .....	108
ALTARES .....	82	COMPLIANCE IT LEGAL CONSULTING.....	110
AMAZON WEB SERVICES .....	84	CONVERTEO .....	112
ATTUNITY .....	86	COUCHBASE .....	114
BEARINGPOINT .....	88	COUTHON CONSEIL.....	116
BLUE DME.....	90	DATASTAX.....	118
BLUESCALE .....	92	DENODO .....	120
BLUESOFT GROUP.....	94	DIGDASH.....	122
BMB SERVICES.....	96	ECS - MEDIASCHOOL EXECUTIVE EDUCATION....	124
BUSINESS & DECISION .....	98	EKIMETRICS .....	126
CENTRALESUPÉLEC EXED.....	100	ELASTIC .....	128
CISCO FRANCE.....	102	EULIDIA.....	130

# Sommaire



**NICOLAS DELATTRE / ORNELLA GODARD**

Feel Data : le Big Data émotionnel

47



**BALA GOPALAKRISHNAN**

Big Data: more than just a fair-weather friend

51



**RAPHAËL CHERRIER**

Mathématiques et prédictif au service de villes durables et intelligentes

54



**ALLAA R. HILAL**

Data for smarter, safer and greener vehicles

57



**JEAN-YVES ROBIN**

Au cœur de la collecte et de l'analyse des données de santé

61



**BENJAMIN CARLIER**

Les startups sportives à l'assaut du Big Data

64



**KÉVIN VITTOZ**

La Data fait son show dans la Billetterie Événementielle

67



**AMÉLIE DELOFFRE**

Quand Data & Gamification entre dans la course

69



**ELIAS HADAYA**

Customer Intelligence et Big Data au service d'une ONG

71

## FOURNISSEURS

74

# BIGDATA

EXALEAD DASSAULT SYSTÈMES .....	132	REPORT ONE .....	174
EXPERIS IT .....	134	SAAGIE .....	176
FINAXYS .....	136	SAS INSTITUTE S.A.S .....	178
GFI INFORMATIQUE .....	138	SCALED RISK .....	180
HEWLETT PACKARD ENTERPRISE .....	140	SCINETIK .....	182
HURENCE .....	142	SEENK LAB .....	184
IDEATRANS .....	144	SENTELIS .....	186
KALANE .....	146	SINEQUA .....	188
KEYRUS .....	148	SKAPÂNÊ .....	190
MAPR TECHNOLOGIES SAS .....	150	SOFT COMPUTING .....	192
MARKLOGIC FRANCE .....	152	SOPRA STERIA .....	194
MATHWORKS .....	154	SPLUNK .....	196
MATLO .....	156	SYNALTIC GROUP .....	198
MICROPOLE .....	158	TABLEAU FRANCE .....	200
MONGO DB .....	160	TALEND .....	202
NEO TECHNOLOGY .....	162	TELECOM PARISTECH / EVOLUTION .....	204
ONEPOINT .....	164	TERADATA .....	206
OUTSCALE .....	166	TIBCO SOFTWARE .....	208
PENTAHO .....	168	YSANCE .....	210
PROXEM .....	170		
QLIK .....	172		



## CHRISTIAN REIMSBACH-KOUNATZE

Economiste & Analyste des politiques numériques / Direction des sciences, technologies et de l'innovation

**OCDE**

*(Organisation de Coopération et de Développement Économiques)*



# Les Enjeux de la **Data-Driven** Economy

**Christian Reimsbach-Kounatze** a rejoint l'OCDE en 2008 et travaille sur les questions relatives à l'économie numérique et aux technologies émergentes. Il a mené des études sur l'impact des TICs sur la société notamment en termes d'emploi et de compétences, et a plus récemment travaillé sur l'économie de la donnée et l'innovation data-driven. Il coordonne maintenant le projet de l'OCDE sur « le rôle de la data sur la croissance économique et le bien-être ». Il rédige notamment des rapports développant des recommandations du conseil de l'OCDE, afin de développer des standards mondiaux et assurer une cohérence entre les politiques des pays membres.

Il était auparavant chercheur en informatique distribuée au sein de SAP sur des thématiques liées aux marchés de l'information, aux systèmes peer-to-peer et à la Business Intelligence, dans le secteur de la finance. Christian Reimsbach-Kounatze a un background à la fois d'informaticien et d'économiste, obtenu au sein du Karlsruhe Institute of Technology (KIT).

L'**OCDE** fournit des standards et des méthodes pour que les différents pays puissent **comparer et mettre en commun les données collectées**

## QUEL EST LE RÔLE DE L'OCDE DANS LE DÉVELOPPEMENT D'UNE ÉCONOMIE DE LA DATA ?

L'OCDE est un forum, **une plateforme**, qui essaye d'amener les différents pays autour de la table, afin de **favoriser les échanges des politiques sur leurs expériences** et de coordonner leurs efforts politiques. L'OCDE fait des recherches, collecte des données empiriques des différents pays membres et les rend comparables, pour favoriser un dialogue entre les pays. Nous étudions par exemple le phénomène récent des cyberattaques, sur lequel il n'y a pas encore de méthode comparable pour en mesurer l'impact. La quantification des risques sur la vie privée est primordiale, mais pas encore possible. En effet, **les réflexions sur les données de vie privée sont parfois dominées par des idéologies, un fort protectionnisme ou libéralisme, et il manque aux États des données objectives pour mener leurs réflexions. Nous fournissons donc des standards et des méthodes pour que les différents pays puissent comparer et mettre en commun les données collectées.** Le manuel Frascati, développé par l'OCDE en Italie dans les années 60 en est un exemple. L'OCDE travaille sur une approche similaire de mesure dans le secteur du numérique et celui de la vie privée et de la sécurité.

Beaucoup de phénomènes numériques sont internationaux, avec de nombreux challenges transfrontaliers, tels que l'échange de données, l'e-commerce, les taxes, la propriété intellectuelle... Il y a donc une véritable nécessité de créer un débat international. Nous ne sommes pas une organisation spécialisée sur une thématique mais nous sommes structurés en différents directorats, qui traitent chacun d'un domaine politique, comme au gouvernement. D'autres organisations internationales sont concentrées sur une spécialité, mais **la numérisation et le sujet du Big Data nécessitent de mobiliser tous les domaines politiques, car cela impacte tous les pans importants de la société : la santé, la concurrence, les taxes, l'éducation...** L'OCDE est donc très bien placée sur ces sujets transversaux.

Les entreprises sont également représentées, c'est une des caractéristiques de l'OCDE. Non seulement, les pays membres sont assis autour de la table, mais **il y a toujours une place pour les représentants du business, les syndicats, la société civile ainsi que les experts techniques du Numérique, les ONG etc.** Ces différents participants ont ainsi la possibilité d'entrer en dialogue avec les différents gouvernements.

## EST-CE QUE LA LOI SUR LA PROTECTION DES DONNÉES DE LA COMMISSION EUROPÉENNE EST SOUTENUE PAR L'OCDE ?

La Commission Européenne est un membre spécial de l'OCDE et elle prend part à ses négociations. En juin 2013, pendant la finalisation de la réglementation sur les données personnelles, l'OCDE était en cours de modernisation de la recommandation sur les données personnelles. Il y a une véritable influence de l'OCDE sur la Commission Européenne et inversement. Nous travaillons en étroite collaboration mais en notre sein, nous prenons en compte les opinions d'autres pays comme le Japon, la Corée, les Etats Unis, le Canada, le Mexique... L'OCDE n'a pas de position sur la réglemen-

tation européenne. Les Etats Unis se sont posé des questions sur l'impact de cette réglementation sur leur économie et sur leurs relations avec l'Europe dans le cadre du Privacy Shield. La position ainsi partagée par les pays membres de l'OCDE est que **« la réglementation ne devrait pas nuire aux échanges des données et au fonctionnement de l'Internet »**

**La réglementation européenne ne devrait pas nuire aux échanges des données et au fonctionnement de l'Internet**

## VOUS PARLEZ DE DATAFICATION DE L'ÉCONOMIE : SUR QUELS CRITÈRES L'OBSERVEZ-VOUS ?

Les investissements des entreprises dans les données et les bases de données permettent d'illustrer ce phénomène de Datafication dans les différents pays membres de l'Union Européenne. Les investissements augmentent significativement et les taux de données échangées entre les entreprises également, plus spécifiquement les flux de données qui traversent Internet augmentent et non unique-

ment les vidéos, même si ce sont les flux les plus échangés entre les entreprises. Une autre dimension observable est la proportion des Data Specialists dans l'entreprise. **En partant d'une base 100 en 1999, le taux est passé à 150% aux Etats-Unis en 2013.** Cette augmentation de la part de Data Specialists est constatée dans tous les pays et **un risque de pénurie dans l'emploi est observable, même aux Etats Unis, alors que la Silicon Valley est la région du monde qui attire le plus les Data Specialists et Data Scientists.** On se trouve dans une véritable situation de brain drain et c'est un problème auquel l'Europe doit faire face mais également les pays en voie de développement, afin de retenir les jeunes talents qui sont capables de gérer les données d'une manière sophistiquée.

## EST-CE QUE VOUS OBSERVEZ DES TENDANCES FORTES, DES DISPARITÉS DANS CES TENDANCES AU SEIN DE L'EUROPE ?

Nous observons des disparités qui reflètent naturellement le niveau de développement des différents pays. Le Chili par exemple vient d'accéder à l'OCDE et leur politique et niveau de développement dans le domaine des NTIC sont naturellement moins avancés. Cependant, certains pays ont un discours beaucoup plus avancé en comparaison de leur développement économique. **Pour exemple la Colombie qui est en processus d'accession à l'OCDE, développe une politique numérique particulièrement impressionnante, dont les pays de l'OCDE ne peuvent que s'inspirer.** Elle promeut l'utilisation du numérique dans des régions éloignées et encourage fortement la transformation numérique dans les entreprises traditionnelles. La Colombie travaille notamment sur des partenariats entre des multinationales et des petites entreprises pour encourager l'intégration de l'ensemble des processus et de la chaîne de valeur et mutualiser les investissements.





On observe des **incentives** comparables en Corée, qui fait partie des pays les plus avancées de l'OCDE, pour favoriser la coopération entre les petites entreprises et les multinationales, notamment en matière de numérique. 17 clusters de recherche ont été mis en place autour d'entreprises multinationales, telles que Samsung, Hondaï ou encore LG, sur un sujet clé dans chacune des entreprises, qui va être traité avec les TPE-PME dans chaque région. De septembre 2014 à fin 2015, les centres ont supporté environ 600 start-ups et les revenus de celles-ci ont atteint 34 Milliards de Korean Wang en 2015 et créé plus de 280 emplois. L'objectif de ce projet était d'augmenter l'exportation de la Corée, et certaines de ces startups ont un impact désormais reconnu à l'étranger. Pour exemple, Tegway, qui a développé des objets connectés générant de l'énergie sur la base des différences de température, a reçu le Top 10 Grand Prix Awards dans les technologies émergentes reconnues par l'UNESCO.

De retour en France, l'Internet des objets est un domaine dans lequel elle accuse un vrai retard. **40% des brevets déposés dans le monde dans le domaine des technologies numériques étaient liés à l'IoT en 2014.** En 2005, les brevets liés à l'IoT en France représentaient 10% contre 5 à 6% en 2014. La France perd son leadership et sa capacité d'innovation dans l'IoT. La Corée ainsi que la Chine prennent le lead. Cela doit véritablement alarmer les politiciens et l'industrie française. Dans le dernier plan numérique de la France, l'accent est mis sur la promotion de la R&D dans l'Internet des objets. Des mesures sont prises en amont sur le financement de la R&D, mais également pour le soutien des entreprises via des allègements de taxation des entreprises qui font de la recherche. En revanche, une des grandes questions restantes concerne le soutien aux petites entreprises. La plupart des mesures ont davantage bénéficié aux grandes entreprises qu'à celles de petite taille. Par exemple la compensation des dépenses de R&D par une réduction des taxes est peu intéressante pour les TPE et PME, qui ont moins de charges d'imposition que les grandes entreprises.

**IoT 40%**

**40% des brevets déposés dans le monde dans le domaine des technologies numériques étaient liés à l'IoT en 2014**





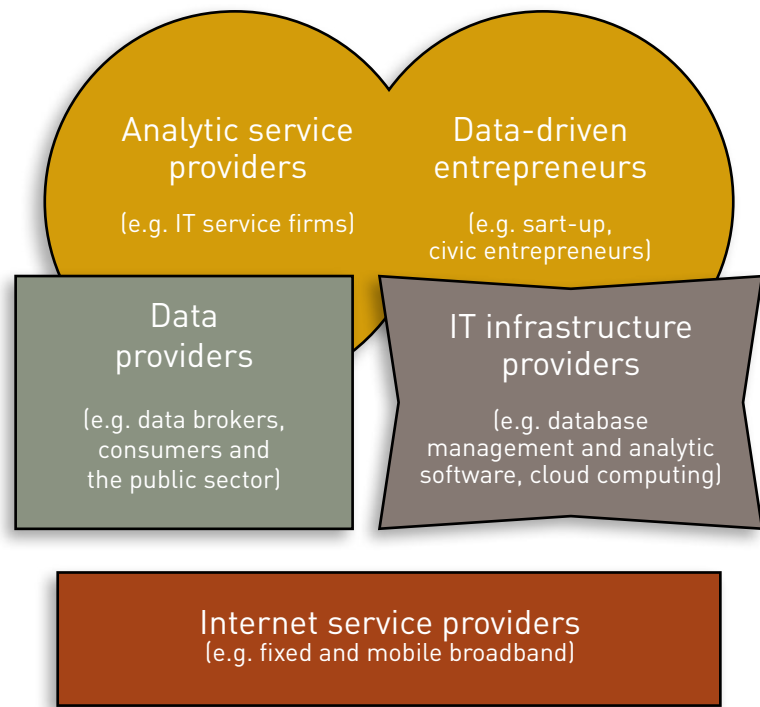
**QUELLES ÉVOLUTIONS OBSERVEZ-VOUS SUR LE PAYSAGE BIG DATA ET LES TYPES D'ACTEURS EN EUROPE ?**

Le Big Data Landscape 2016 est centré autour de la donnée et ne nous semble pas forcément utile pour les décideurs politiques, car il est essentiellement concentré sur les produits des entreprises et leurs services. De plus, les acteurs des télécommunications sont manquants dans le Big Data Landscape, pourtant ce sont eux qui fournissent l'infrastructure primaire pour que les entreprises puissent se connecter au cloud et aux réseaux, et leur activité leur permet également de collecter des données pour développer des services. Orange collecte par exemple les données des téléphones mobiles pour comprendre les flux de personnes et les fournit aux entreprises qui proposent des systèmes de navigation.

**Les Télécoms agissent comme Data Providers** et TomTom est par exemple dans la position de monétiser ses données grâce aux acteurs des télécommunications.

Le Big Data, selon nous, ne représente qu'un coût d'utilisation. **C'est l'innovation issue du Big Data et des données qui doit être mise en avant.** L'Analytics est par exemple augmentée

**The data ecosystem as layers of key roles of actors**



par l'intégration des données, il faut ainsi toujours observer les données avec les softwares associés. Nous proposons donc une visualisation plus simple des acteurs d'après leurs rôles pour le politique, pas concentré sur la technique et les logiciels. (cf : The data ecosystem as layers of key roles of actors)

Le potentiel des marchés dans les couches basses de l'écosystème data est plutôt faible car les acteurs y sont déjà établis. Les opportunités de nouveaux marchés diminuent, il y aura donc très peu de startups dans les télécommunications et peu de nouvelles offres d'infrastructures Cloud, Data Base.

En revanche, davantage d'entreprises fleurissent sur le marché des **Data Providers**, spécialisés dans la collecte, l'analyse des corrélations et les combinaisons avec de la visualisation.

Les **Data Providers** ou fournisseurs de données vont offrir aux consommateurs des services complémentaires par exemple en développant des applications mobiles.

Et dans le secteur de l'**Analytics**, les services proposés vont être de plus en plus personnalisés et adaptés aux besoins de l'entreprise client, la plupart réalisant des missions de conseils dans le BtoB.

Ces deux derniers types d'acteurs sont les plus dynamiques et ceux dans lesquels les startups sont les plus nombreuses.

**L'internet des objets va également être un véritable catalyseur en augmentant la capacité à générer des données.**

**De nouvelles opportunités pour le BtoB vont voir le jour mais également directement pour les consommateurs via les wearables.**

**QUELS SONT LES GRANDS ENJEUX DE L'ÉCOSYSTÈME DATA ET LES CHALLENGES À ADRESSER ?**

**L'ÉCOSYSTÈME DES APIS**

Le premier enjeu dans l'écosystème Data réside dans le cytoplasme. L'écosystème data est comme une cellule dans laquelle les différents éléments vont interagir. Les APIs sont le liant entre les différents acteurs data, elles seront celles qui vont contrôler et dominer l'écosystème. Il est ainsi impératif de réguler leur utilisation. **Une bataille légale autour des droits intellectuels liés aux APIs a vu le jour.** Oracle contrôle les APIs fondées sur Java, dont il est propriétaire. Android dispose d'un logiciel très puissant utilisant Java. Oracle a donc essayé de capitaliser sur Java pour tirer des bénéfices et contrôler le développement d'Android mais Google a démontré que Java était en code ouvert, utilisable librement par tous les développeurs et que son utilisation avait été raisonnable dans le développement d'Android. Oracle a donc perdu en juin 2016 son procès contre Google mais l'écosystème des APIs reste à régler.

De même, pour accéder à Twitter et pour construire des applications autour de Twitter, il faut avoir accès aux APIs. La réglementation et gouvernance des APIs de Twitter n'est pas claire et les modalités d'accès changent fréquemment. Il y a donc une insécurité pour les entreprises qui souhaitent accéder aux APIs de Twitter et une obligation pour eux de renégocier cet accès régulièrement. Il ne s'agit pas forcément d'un problème d'ouverture mais de standard qui doit être connu, approuvé et en vigueur dans l'écosystème des APIs.

### LAGRÉGATION DES DONNÉES EN QUESTION

Une discussion est ouverte à Bruxelles sur le cas de Google pour savoir si le fait d'agréger des données ouvre des barrières de concurrence sur le marché. **La position de l'OCDE est la suivante, si l'accès à un marché dépend de l'accès aux données, les données agrégées par les entreprises qui ne peuvent être générées par d'autres entreprises de la même manière et en même quantité, devraient attirer l'attention de l'autorité de la concurrence.** Jusqu'à présent les atteintes à la vie privée à travers les données personnelles n'étaient pas considérées comme une violation par l'Autorité de la Concurrence. Pour qu'elle puisse intervenir, il faut qu'elle ait des preuves qu'il y a eu violation d'une position dominante en imposant à ses clients des conditions inéquitables ou excessives.

Le consommateur est clé pour analyser si la dominance de marché existe véritablement. C'est toujours le prix qui est observé et son augmentation qui témoigne de l'abus de position dominante. Or, dans le numérique, la plupart des services sont offerts ou sans prix. Il faut donc déterminer quel est le facteur qui témoigne de cet abus. **Une recommandation de l'OCDE est de se baser sur la violation de la vie privée. Les autorités de la concurrence doivent travailler ensemble avec les autorités de la vie privée et les consommateurs.**

### CLOUD COMPUTING

**40% des grandes entreprises utilisent le Cloud Computing.** Les technologies Cloud sont également clés pour les petites entreprises, dans la mesure où elles ne nécessitent pas d'investissements massifs, même si à l'heure actuelle seulement 10% ont investi dans le Cloud. L'interopérabilité des services, mais également la portabilité des données d'un service à l'autre, doivent donc être assurées, dans un contexte où les acteurs des NTIC ont essayé de monétiser leurs services numériques, notamment Cloud, en s'appuyant sur le phénomène de lock in. Axelle Lemaire soutient la portabilité, ce qui est plus que bienvenu.

### OPEN DATA ET CONFIANCE

**La confiance est au cœur de l'écosystème,** c'est une étape intermédiaire indispensable. L'Open Data ne doit pas signifier impérativement l'ouverture extrême des données, mais avant tout l'échange des données des entreprises dans le réseau, avec les acteurs qui partagent les mêmes intérêts. Réinstaurer la confiance pour l'ouverture des données est indispensable. Le respect de la vie privée est souvent perçu comme un frein à l'innovation, mais je pense au contraire qu'il s'agit d'un driver et d'une condition de l'innovation. Beaucoup d'entreprises disent que le respect de la vie privée les empêche d'accéder aux données, mais elles ne sont elles-mêmes pas prêtes à ouvrir leurs données aux concurrents sur leur marché. La problématique de confiance concernant les données de l'entreprise est identique à celle des utilisateurs. **Il faut instaurer une confiance autour de la manipulation qui est faite des données.**

### VIE PRIVÉE, PROPRIÉTÉ DES DONNÉES ET NON-DISCRIMINATION

**La protection de la vie personnelle est naturellement un des enjeux fondamentaux de l'écosystème data, la cybersécurité en est un second, de même que la protection des droits intellectuels :** comment les données sont-elles touchées par l'amplitude de la réglementation ? Quelle régulation est appliquée pour protéger les données non personnelles et qui en est propriétaire ?

Même pour les données personnelles, le principe de propriété n'a pas encore de réponse, il est difficilement applicable avec les réglementations en vigueur. Le « Data base protection » est la seule réglementation implémentée dans les différents pays européens qui protège la collecte des informations. Cette collecte nécessite un effort intellectuel pour agréger les informations et est donc protégée par le régime de la protection intellectuelle. Ainsi, si une base de données contient des sélections de données, alors elle sera protégée par le droit de la propriété intellectuelle. S'il s'agit d'une très grande base de données non travaillée, alors elle ne bénéficiera pas de protection.

La vie privée est un des plus gros challenges. Le risque qui vient avec l'exploitation des données des individus va au-delà des aspects légaux et est très lié à l'éthique. **Les données peuvent être protégées par la loi mais leur utilisation et manipulation peuvent se faire dans un but moins éthique.** Pour exemple, les algorithmes de Target, utilisés en marketing, permettent d'identifier le mois du début de la grossesse. On joue ici sur un moment de vulnérabilité, caractérisé par des changements de préférence à une période précise. Cela rentre dans le champ de la manipulation et de l'exploitation d'une position dans laquelle l'entreprise détient plus d'informations que l'utilisateur.

Enfin il faut adresser un **challenge social relatif à la non-discrimination.** Les algorithmes, par exemple, mettent l'accent sur les stéréotypes, ils augmentent l'efficacité du Profiling mais peuvent mener à des injustices car ils renforcent les facteurs dominants. De même le traitement différent des personnes en fonction des risques qu'elles encourent, par exemple en matière d'assurance santé va peu à peu remettre en cause le concept de mutualisation des risques.

Toutes ces questions vont être traitées par l'OCDE en 2017 et **la France et le Royaume-Uni sont parmi les pays leaders dans la réflexion sur la gouvernance des données,** notamment via le projet mydata, initié en 2011, dans le but de permettre aux consommateurs d'accéder aux données que les sociétés privées possèdent sur eux afin de mieux orienter leurs choix. **La France mène également le débat sur les données d'intérêt public, tout comme la Norvège, la Suède et le Danemark qui sont aussi en réflexions avancées sur la gouvernance des données.**





# MÁRTA NAGY-ROTHENGASS

Head of Unit «Data Policy and Innovation»

DG Connect

**EU COMMISSION**



## The framework conditions for the European Data Economy

**Marta Nagy-Rothengass** has a broad working experience Europe wide in business and public environment, and gathered her interest on ICT while establishing «New Media» by a traditional German manufacturing company. In 2005, she joined the EC as the Head of Unit of «ICT for the Environment», further developed as «ICT for Sustainable Growth». After moving to Luxembourg in 2008, she served as Head of Unit «Technologies for Information Management».

Since July 2012, she is in charge of the Data Policy and Innovation Unit (formerly known as Data Value Chain Unit) and implements a strategy to extract the maximum value from data by building on the intelligent use of data sources across Europe and beyond.

The European data market amounted to more than EUR 47 billion in 2013, raising to EUR 50.4 billion in 2014, at a growth rate of 6.3%

## OVERVIEW OF THE EU DATA MARKET

I can see that over the last years, data have become a key factor of production across the economy. It has become a tradable good with major socio-economic value. We have commissioned a study to set up a European Data Market Monitoring Tool which provides some key figures. Let me share some of them with you:

**The European data market amounted to more than EUR 47 billion in 2013, raising to EUR 50.4 billion in 2014, at a growth rate of 6.3%.**

This represents a share of total ICT spending in the EU28 of 8.7% in 2014, which is significant for an emerging market.

The overall value of the data economy in the EU is estimated at about EUR 255 billion in 2014, representing a contribution to the EU GDP of approximately 1.8%

**The overall value of the data economy in the EU is estimated at about EUR 255 billion in 2014, representing a contribution to the EU GDP of approximately 1.8%.** The value of the data economy includes the estimates of all the economic impacts produced by the adoption of data-driven innovation and data technologies in the EU: this comprises direct impacts of the data industry and its suppliers, indirect impacts on user industries, and induced impacts created by the additional growth and spending generated by data-driven innovation across the whole of the European economy.

In 2014, the European data industry comprised approximately **243,000 companies with a share of 14% of the 1.7 million enterprises** populating the ICT and professional services sectors. This includes start-ups, innovative SMEs, and many existing enterprises that are moving to exploit the emerging business opportunities of the data market. Data companies' revenues amounted to almost EUR 48 billion in 2013 and EUR 51 billion in 2014, thus registering a **considerable growth rate of 7.1% year-on-year**. I should underline the fact that we see **in 2014 a dynamic data industry, but with a still immature data user population**.

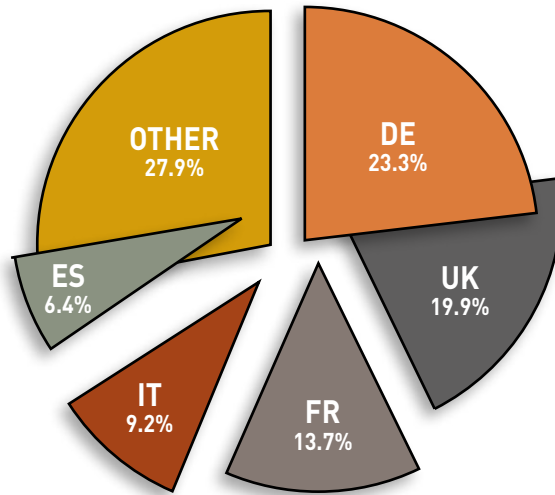
I strongly invite you to have a look at the website developed in the framework of this study on EU Data markets (<http://www.datalandscape.eu/>).



## ARE THERE NATIONAL DISPARITIES IN THE EUROPEAN LANDSCAPE?

**We see that Germany, the U.K., France, Italy, Poland, and Spain («the Big Six») accounted for more than 70% of all data workers in the European Union in 2014.**

### Data markets in the Member States of the EU



(Source: European Data Market Monitoring Tool, IDC 2015)

In terms of employment share, though, the discrepancy between large economies and smaller economies tends to lose importance and appears to be more correlated to other variables such as the overall ICT penetration rate by country: this is why Member States like Luxembourg and Lithuania, for example, have data worker shares well above the EU average. On the other hand, countries like Slovakia and Romania still face great challenges from the point of view of data worker share.

## HOW IS FRANCE POSITIONING ITSELF IN THIS GLOBAL EU DATA MARKET?

Looking at data companies share of total companies by Member States in 2014, we see that France has a sizable number of data companies but a large proportion of small, traditional companies that may take time before they take up data technologies.

With its Loi du Numérique, but also with a number of actions under its industrial policy «Nouvelle France Industrielle - Industrie du Futur», France is setting the standard on a number of issues important for the on-going digital revolution. The innovative thinking of the Conseil National du Numérique, Teralab and Etalab in the area of open data and big data underline the commitment of France to drive the agenda on data-driven innovation in Europe.

## THE MILESTONES OF THE COMMISSION'S ACTIONS FOR A EUROPEAN DATA ECONOMY

As you might remember, the October 2013 European Council concluded that: «EU action should provide the right framework conditions for a single market for big data». In response, in July 2014, the Commission outlined a new strategy on data-driven innovation, supporting and accelerating the transition towards a data-driven economy in Europe.

Framework conditions can be improved with better legal certainty around the data concept. Therefore, the Commission also announced in the Digital Single Market Communication that it will propose in 2016 a **European «Free flow of data» initiative that tackles restrictions on the free movement of data for reasons other than the protection of personal data** within the EU and unjustified restrictions on the location of data for storage or processing purposes. It will address the emerging issues of ownership, interoperability, usability and access to data in situations such as business-to-business, business-to-consumer, machine generated and machine-to-machine data. We will encourage access to public data to help drive innovation. We see big benefits in opening up governmental data for re-use for citizens, businesses, and society and for the governments themselves. In this context, the Commission has launched a fully-fledged pan-European infrastructure to federate content published on European public open data portals through one single interface. The European Data portal is operational since November 2015, currently providing a unique access to more than 400 000 datasets from all over Europe.

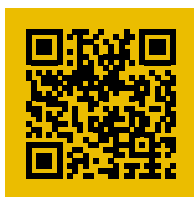
In order to boost investment and foster community building the Big Data value, a **contractual Public-Private Partnership on data was created in October 2014** with the concrete objective to fund "game-changing" data innovation ideas, building on a Strategic Research and Innovation Agenda. European Industry is represented by the **Big Data Value Association** (grouping over 130 members). Investments into Research and Innovation are expected to reach around **EUR 2.5 billion over 2016-**

2020 (of which approximately EUR 500 million from Horizon 2020 programme). Evaluation of the first round of proposals is now under way.

In terms of support to collaborative research and innovation actions, you must have come across - maybe without knowing it is actually supported by the EU - some of the more than 100 projects funded by EU research and innovation programs like FP7, CIP and H2020 addressing different data intensive sectors such as energy, transport and health.

### THE OBJECTIVES OF THE STARTUP PROGRAM HORIZON 2020

Startup Europe is indeed a key program that aims at strengthening the business environment for web and ICT entrepreneurs so that their ideas and business can start and grow in the EU. Since 2011, it has developed a portfolio of actions focused on developing a connected continent of startup ecosystems. We believe in the importance of a developed ecosystem to nurture startups from early stage up to growth and expansion stages. I invite you to have a look at the One Stop Shop ([www.startupeuropeclub.eu](http://www.startupeuropeclub.eu)), that has been developed to ease the access to all the services provided to startups and ecosystem builders.



### WHAT ARE THE PRIORITIES FOR 2017?

Important proposals will come in 2016 and 2017 as part of the Digital Single Market which is as you know among the top priority of our Commission. There are still many challenges to address. We need to **facilitate access to digital computing infrastructures**: for industry, especially for SMEs, and for research centres. We need new approaches supporting the emerging **data economy ensuring access to good quality data, e.g. exchange of data as commercial objects**. We also need a **skilled workforce able to contribute to and benefit from the digital transformation**.

**Legal certainty is a very important element**. This is why, as we already mentioned, we are going to address the issues of data **«ownership», interoperability, (re)usability and access to data, and liability**. These emerging issues (including data «ownership») still require substantial work and are still subject to consultations and assessment; no decisions have been taken about the nature of any proposals, but they will keep us busy this year and the next.

I also think it is high time we tackled restrictions on the free movement of data for reasons other than the protection of personal data within the EU and unjustified restrictions on the location of data for storage or processing purposes. In the coming months and years, we will continue our cooperation with the European industry, since only if the business community joins forces then digitisation of industry can unleash its full potential.

The support to innovation is paramount and remains our focus. As an example, I would like to mention **the Open Data Incubator called ODInE**, a EUR 7.8 million project (Horizon 2020 funding) that became operational in spring 2015. It is supporting up to 50 start-ups to experiment with small grants allowing them to create sustainable commercial applications built on open data.

As an example of the already achieved results of this action, **CommoPrices, a French startup supported by ODInE**, enables users to easily track the price of any commodity, on a single portal. In this context, ODInE contributed to the validation and the acceleration of the business, the improvement and broadening of the data, and the company's expansion into other countries. This is the kind of support needed by rising small businesses in Europe.

### WHAT IS YOUR OPINION ON THE SKILLS SHORTAGE FOR DATA SPECIALISTS?

Skills shortage can be explained by the complexity of the different skills needed to understand data, sectors and decisions related to it. We don't only see new disciplines appearing, but also professionals in companies needing to enlarge their skills in the field.

**The European Union had 5.7 million data workers in 2013 and 6.1 million data workers in**

**2014, representing more than 3% of the total employment in the EU.** The number of data workers therefore increased by almost 6% year-on-year in 2014, which is well over the 2.2% growth rate for employment in the EU registered in 2014. This shows that data workers are actually more than usually thought! Still, the main potential gap is for data scientists and «hard» technology skills. And **I see the shortage of skills in the field of data as one of the major threats to the development of this sector in Europe.** The right skillset is essential to properly process data and create new businesses.

The European Union had **5.7 million data workers in 2013** and **6.1 million data workers in 2014**, representing more than 3% of the total employment in the EU



With the DSM Strategy, the Commission has committed to address digital skills and expertise as a key component of its future initiatives on skills and training. Last year, we set up the European Data Science Academy (EDSA). This is a structure that develops learning materials based on an analysis of the need for skilled data workers and the professionals' needs in terms of trainings. As you may know, **the Commission published on the 10th of June 2016 a new Skills Agenda for Europe, working together to strengthen human capital, employability and competitiveness.** It covers a number of actions and initiatives with the ambition to tackle the digital skills deficit in Europe by improving the quality and relevance of skills formation, making skills and qualifications more visible and comparable, and advancing skills intelligence, documentation and informed career choices. Finally, through the European Research and Innovation Programme Horizon 2020, the Commission will continue to promote data skills by supporting activities like the alignment of curricula and training programmes to industry needs, the establishment of national centres of excellence in all Member states, and exchanges of students and data professionals across Member States.

#### *HOW WILL COMPANIES BEST BENEFIT FROM A EUROPEAN DATA ECONOMY?*

Based on a continuous growth scenario, **the sector is expected to show a healthy growth, with a European data market reaching EUR 83 billion.** It is also expected to bring an improved efficiency in various sectors, with new and more targeted services, the reuse of data in a cross-sectorial way, and the development of platforms (for instance, industrial data spaces).

The potential is more than big. We see that data-driven innovation is taking place in many contexts, in manufacturing and the services industry, but will soon also encompass every as-

pect of our life. Just generating the data, however, will not ensure that the data can always be used in a way that is efficient in economic and societal terms.

Let me give you some examples: The connected car of the future is not only a consumer of data, but also a producer of data, e.g. on weather and road conditions. The raw data collected by the car can be fed back and can be used in order to enhance weather prediction, ever more granular and in real-time. It goes the same for airplane engines, wind turbines and smart meters. In the near future, an increasing number of intelligent devices and components that are part of our daily lives will be data users and data producers.

This creation of value chains is key to keep our industries competitive. Competitiveness will come from the ability of industries to provide additional data services on top of the products they sell. This happens in the machine tool business already, where leading companies do not sell you the machine only; they sell you a service that uses this machine and monitors it to optimise run-time, to eliminate break-downs and to reduce overall costs of use.

#### *WHAT IMPACT OF THE 14 APRIL 2016 REGULATION AND DIRECTIVE ADOPTED BY THE EUROPEAN PARLIAMENT?*

Clearly, the EU Data Protection Reform is an enabler for data-driven services in Europe. Consumers need to trust companies in order to take up the services they offer. So in this respect, privacy-friendly companies have a competitive edge, and I believe the privacy-friendly environment in Europe is one of the incentives that can bring innovative technology companies to set up shop in the EU. This is illustrated by big companies investing in European Data centres to develop cloud solutions. I would say the GDPR reflects a certain societal

consensus in Europe on what should be possible with personal information and is thus crucial for the acceptance of data analytics. It also decreases compliance costs overall by setting one single legal framework for the entire EU and streamlining administrative decision-making for data processing under the «one-stop-shop» principle.

*[The views expressed in the article are the sole responsibility of the author and in no way represent the view of the European Commission and its services.]*

The **European data** market is expected to reach **83 billion EUR**



# ALAIN BENSOUSSAN

Avocat à la Cour d'Appel de Paris

ALAIN BENSOUSSAN AVOCATS LEXING



## Chronique juridique du Big Data, des algorithmes et de l'Intelligence Artificielle



Avocat à la Cour d'appel de Paris, précurseur du droit des technologies avancées, **Alain Bensoussan** a fait de l'élaboration de concepts nouveaux l'une de ses marques de fabrique : domicile virtuel, droits de l'homme numérique, vie privée résiduelle... Il fonde et préside le cabinet Alain Bensoussan Avocats Lexing qui comprend aujourd'hui plus de 120 collaborateurs dont environ 80 avocats. En 2012, après avoir créé Lexing®, premier réseau international d'avocats technologues, il lance ainsi au sein de son cabinet un département de droit des robots, y voyant « la reconnaissance par le droit d'une mutation technologique au moins aussi importante que l'ont été l'informatique et les réseaux sociaux au 20<sup>e</sup> siècle ». Aux yeux de cet infatigable explorateur de nouveaux domaines du monde numérique, il était temps de créer un droit des robots les dotant d'une personnalité et d'une identité juridique pour en faire, demain, des sujets de droit : « Avec l'introduction d'une Intelligence Artificielle, les robots ne sont pas de simples automates. Ils ont des capacités grandissantes qui les amènent à collaborer avec les hommes.



## QUELLE RÉALITÉ REVÊT LE BIG DATA ?

Le Big Data stricto sensu, pour le juriste, regroupe les Data, mais également les algorithmes utilisés afin de tirer un résultat de celles-ci. **Les Data sont en quelque sorte des données en état futur d'interprétation. Les Data sans algorithmes sont un peu comme des violons sans archet.** Le Big Data est aujourd'hui capable de grandes prouesses par la multiplication de ces analyses, qu'elles soient statistiques ou liées à des technologies d'Intelligence Artificielle. Lorsque l'on combine Data et algorithmes, on obtient la possibilité de créer un double informationnel de l'individu. Dans ce cadre, il faut disposer d'une très grande quantité de données, que l'on appelle des « lacs de données » composés de données de formats multiples, qui énoncent des signaux faibles interprétés par les algorithmes de manière intelligente, pour déduire des prédictions à partir de cohortes de comportements.

...

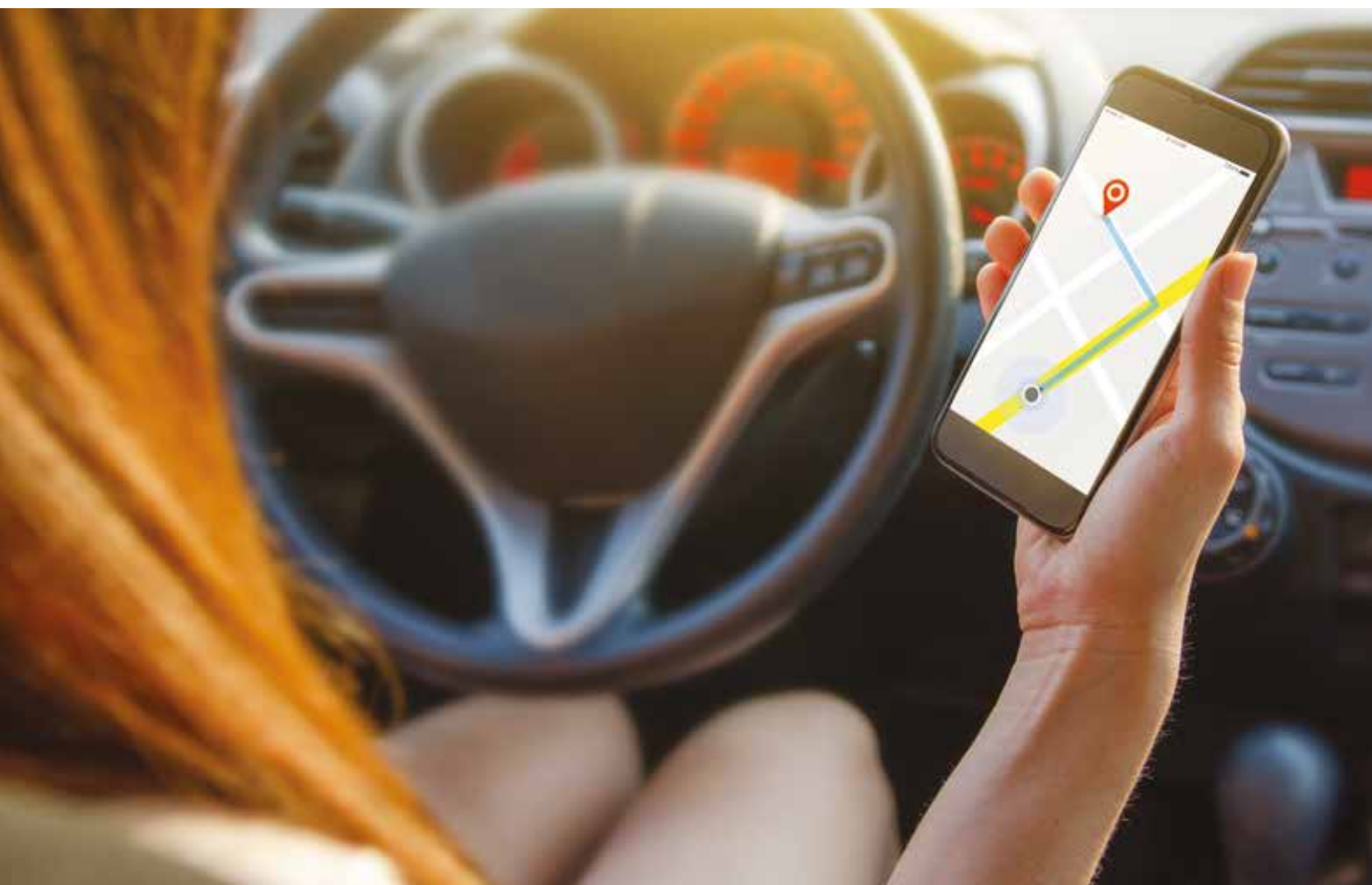
Les **Data sans algorithmes** sont un peu comme des **violons sans archet**

Les algorithmes induisent ainsi que les gens qui aiment le jazz seront « condamnés » à recevoir un pop-up toute leur vie leur recommandant, avec une probabilité d'exactitude supérieure à 90%, les derniers romans policiers à la mode. Nous sommes donc sous contrainte musicale et littéraire algorithmique ! Le monde de demain sera piloté par les algorithmes. D'où l'importance d'analyser leurs risques et de comprendre ce qu'ils font.

## QUID DE LA SÉCURISATION DE CES ALGORITHMES ?

Un des principaux enjeux du Big Data est justement celui de la gouvernance des algorithmes. Il est nécessaire de créer un droit des algorithmes avec :

- **une obligation d'information, permettant de savoir ce que fait l'algorithme ;**
- **une obligation éthique pour que les algorithmes soient conçus « ethics by design » et ne dépassent pas certaines limites d'intimité ;**



... Ainsi, les algorithmes Predpol (Predicting Policing Software), par exemple, ont notamment permis de faire baisser le taux de sinistralité de manière très importante à Los Angeles. De même, la brigade des pompiers de Londres utilise ce type d'algorithmes pour prévenir des zones à risques dans lesquelles, par exemple, un incendie pourrait se déclencher.

Les algorithmes sont utilisés pour optimiser tous les comportements des individus face à des opérations, qu'elles soient d'intelligence, de déplacement, de consommation... **Nous allons vers une société qui dès 2020 sera entièrement pilotée par les algorithmes.** Ce sera le

début de « l'esclavagisme technologique » que l'on peut illustrer de la façon suivante : **nous serons téléguidés en matière de transports, télésurveillés en matière de sécurité, télésuggérés en matière de culture, téléopérés en matière de santé et enfin téléorientés en matière de commerce !** Nous le sommes d'ailleurs déjà : nous nous abandonnons dans les bras des algorithmes. Par exemple, lorsque nous entrons une adresse dans un GPS ou un ordinateur de bord, nous abandonnons la direction à l'algorithme ; parfois même nous lui parlons et nous profitons de sa compétence dans la mesure où l'algorithme GPS prend en compte les embouteillages, les conséquences des accidents et même nos préférences, si elles sont enregistrées.

- **la création d'un commissaire aux algorithmes** chargé de vérifier l'adéquation et la conformité des algorithmes aux règles définies pour les plateformes, le consommateur étant dans l'impossibilité de les contrôler lui-même.

Les consommateurs sont en position anxieuse s'agissant de leurs données à caractère personnel mais ils les partagent assez facilement car dans ce partage, ils obtiennent des droits qui pour l'instant les intéressent. Une société comme Facebook, forte de ses 1,6 milliards d'utilisateurs, répond à l'évidence à un besoin, qui lui-même relève de l'essence même des hommes et des femmes, à savoir



communiquer.

J'appelle ainsi de mes vœux un nouveau corps de consultants ainsi qu'un commissaire aux algorithmes, qui, à l'instar du commissaire aux comptes, agirait pour le compte des dirigeants et associés, des salariés, des clients... Il serait l'homme permettant aux entreprises de se situer dans un état de conformité algorithmique. Ceci pour parer à tout risque systémique. Les algorithmes ne peuvent pas être totalement laissés à une régulation de marché.

### QUELS SONT LES PRINCIPAUX POINTS DE LA RÉFORME EUROPÉENNE SUR LA PROTECTION DES DONNÉES PERSONNELLES QUI VIENT D'INTERVENIR ?

La protection des données personnelles en Europe va être modifiée de manière radicale par le **règlement sur la Protection des données à caractère personnel 2016/679 du 27 avril 2016**, qui entrera en vigueur en France le 25 mai 2018.

Ce nouveau système juridique reprend une partie ancienne de la réglementation sur la protection des données. On y retrouve les dispositions relatives au :

- **Droit d'information,**
- **Droit des données à caractère personnel,**
- **Droit des traitements.**

Le tout devant être fait avec exactitude, pertinence, en respectant les finalités spécifiques. Or, c'est précisément ce dernier point qui pose problème car dans le Big Data, il n'y a pas essence pas de finalités spécifiques. Il existe des finalités d'orientation, des grands axes en quelque sorte, mais la collecte initiale n'a pas pour objet d'être respectée à l'identité. On cherche dans les lacs des données à mélanger des données d'origines différentes, obtenues dans le cadre de finalités différentes, afin de faire parler les données et de leur donner du sens à partir des signaux faibles.

Il s'agit d'une des préoccupations importantes pour les professionnels de la Data. Il faut mettre en place une règle de gouvernance au regard de cet impératif de finalité de la réglementation.

Par ailleurs, dans la pratique, la plupart des entreprises, notamment dans le domaine du Big Data, ne respectaient pas l'ancienne réglementation. Notamment parce que les sanctions encourues n'étaient pas très dissuasives puisqu'elles étaient plafonnées à 150 000 euros et 300 000 euros en cas de récidive, sous réserve de condamnation pénale. Aujourd'hui, **dans le nouveau « paquet » réglementaire, le risque maximal est de 20 millions d'euros ou 4% du chiffre d'affaires !** On change complètement d'échelle.

On trouve trois grandes obligations complémentaires incluses dans la réforme de la protection des données personnelles :

#### LA PROTECTION DÈS LA CONCEPTION

**Le Privacy by design ou plus précisément « protection dès la conception »** signifie qu'avant même de coder, on doit réfléchir à la mise en place de la protection des données dans le cadre d'une collecte licite, loyale, également dans le cadre d'un traitement à finalité spécifique, respectueux des données

sensibles. Ces exigences sont à prendre en compte dès la phase de conception, de réalisation, de démarrage, d'exploitation, et ce jusqu'à la phase de maintenance. On doit penser en amont à la durée de conservation des données nécessairement limitée, avec dans le cas général des données non nominatives (les données nominatives devant être l'exception), qui seront obtenues uniquement avec un consentement. **Au terme de la finalité, il faudra détruire les données.**

#### LA PROTECTION PAR DÉFAUT

La conformité est par nature instable en raison des innovations technologiques. Le Règlement Data Protection prévoit donc que le responsable doit garantir que la conformité est bien respectée et prendre toutes les mesures nécessaires en cas d'écart pour une remise en conformité et pour stabiliser juridiquement ses innovations.

#### L'ACCOUNTABILITY

Ce terme peut être traduit par « responsabilité » mais pas au sens de responsabilité pour faute ou sans faute. **Il faut que la personne soit en mesure de documenter l'ensemble des éléments de conformité et qu'elle puisse rapporter la preuve de celle-ci.** C'est au responsable de traitement des données qu'il revient de prouver qu'il est en conformité et non à l'autorité de régulation de démontrer qu'il y a non-conformité.

La réglementation européenne spécifie également l'obligation :

- de procéder à des études d'impact lors du traitement de données à risque,
- de prévenir, en matière de faille de sécurité des données, les autorités dans un délai de 72h maximum,
- de tenir un registre des failles de sécurité : les entreprises doivent communiquer à la première demande des autorités de régulation sur les failles de sécurité, et informer les personnes concernées des conséquences entraînées par cette faille de sécurité.

### QUELS SONT LES IMPACTS CONCRETS DU RÈGLEMENT EUROPÉEN SUR LES ENTREPRISES ?

Les conséquences budgétaires pour les entreprises sont très importantes, notamment car elles doivent prévoir de réaliser un audit de conformité, de situation, et mettre en place un plan de conformité pour régulariser leur situation. A cela s'ajoute la nécessité de mettre en place une politique générale de conformité après la phase de régularisation pour un maintien en conditions opérationnelles.

### QUELLES SONT LES LIMITES DU RÈGLEMENT ?

La véritable limite du Règlement Data Protection réside dans la prise en compte du Big Data et de l'Intelligence Artificielle. **Le Règlement concerne les technologies numériques et est fortement orienté « silos d'information ».** Le Big Data est donc violemment impacté par deux des principes du Règlement, celui de finalité spécifique déjà évoqué, et celui de minimisation des données.

Le principe de minimisation figure entre parenthèse dans le Règlement et n'est donc pas fa-

cile à interpréter. **La minimisation signifie que le responsable de traitement doit rechercher des solutions permettant de délier l'identité de la personne des données la concernant et d'effectuer des traitements de données à condition qu'elles ne soient pas à caractère personnel.**

Ce principe n'existe pas aux Etats Unis. Les données nominatives ne sont pas sous contrainte de minimisation. **C'est un handicap pour l'Europe car ce principe est contraire aux fondements du Data, du Machine Learning et du Deep Learning,** dont le principe est justement la maximisation et l'augmentation de la taille du lac de données pour permettre aux algorithmes de traiter plus de données et aux machines d'être plus opérationnelles. Le principe de minimisation se présente de manière assez orthogonale par rapport aux algorithmes et **il est ainsi nécessaire de repenser le règlement vers une finalité d'orientation.**

**Le Big Data est violemment impacté par deux des principes du Règlement, celui de finalité spécifique et celui de minimisation des données**

### QUI EST PROPRIÉTAIRE DES DONNÉES ?

C'est une des grandes questions que pose le Big Data. Il n'existe pas de réglementation dans le monde sur la reconnaissance de la propriété des données, bien que celles-ci soient reconues comme une chose et fassent par exemple, à ce titre, l'objet d'une réglementation pour vol. Il s'agit d'un débat entre économistes sur la reconnaissance d'un droit de propriété ou droit d'accès. En ce qui me concerne, je milite justement pour la reconnaissance d'un droit de propriété.

Les contrats en matière de base de données présentent également de grandes difficultés de qualification. Les bases de données ne revêtent que rarement le caractère d'originalité requis. Le producteur de la base de données détient un droit sui generis qui protège le contenu de celle-ci, mais les données ne représentent pas forcément un investissement important, car elles sont généralement non formatées. **La valeur des bases de données réside dans la réunion des données plus que l'intelligence pour les traiter.**

## QUID DE LA RÉGLEMENTATION EN MATIÈRE D'INTELLIGENCE ARTIFICIELLE ?

Indépendamment du Big Data et des algorithmes, ces outils juridiques peuvent être utilisés au service de nouvelles technologies d'Intelligence Artificielle, les réseaux neuronaux, l'apprentissage profond Deep Learning et l'apprentissage automatique Machine Learning. Je pense qu'il est également nécessaire de redéfinir le régime et système de responsabilité face à ces intelligences faibles. Ces intelligences peuvent modifier le réel. L'exemple type est la voiture intelligente : il est nécessaire de définir si l'utilisateur ou la plateforme d'intelligence sont responsables en cas d'accidents. De même les robots, en fonction des réglages effectués, ne donneront pas systématiquement les mêmes réponses. Il est donc nécessaire de définir si la plateforme ou l'utilisateur est responsable en cas d'accident ou si le propriétaire/fabricant l'est conformément au régime général.

Dans le domaine de la robotique, la nature doit également être définie. **Je milite pour une personnalité juridique propre des robots, la définition d'une nouvelle espèce artificielle accompagnée d'une obligation de traçabilité.** Il faut nécessairement une boîte noire pour assurer une « traçabilité by design » et suivre les déplacements des robots.

Il faut également considérer juridiquement le respect de l'intimité par les robots qui détiennent des informations personnelles sur leurs compagnons, dans la mesure où ils enregistrent tout ce qui se passe dans une pièce et captent des moments de vie privée. Le robot ne doit ni trahir son compagnon en communiquant des informations aux tiers non autorisés, tels que le fabricant, ni prendre de décisions sans l'autorisation de son compagnon.

Les schémas de responsabilité en matière d'Intelligence Artificielle et de robotique doivent donc être repensés. Je milite également pour la création d'un comité d'éthique pour la robotique afin de reconsidérer la dignité des robots et le rapport robot-individu.

**Je milite pour une personnalité juridique propre des robots, la définition d'une nouvelle espèce artificielle accompagnée d'une obligation de traçabilité**





## JÉRÔME BÉRANGER

Co-fondateur et Chief Scientific Officer

**ADEL**

*(Algorithm Data Ethics Label)*



# De l'éthique dans l'exploitation et le traitement des données numériques

Economiste de la santé, **Jérôme Béranger** a travaillé pendant 4 ans comme consultant en gestion de projet, stratégie et management pour des cabinets de conseil spécialisés en santé, hospitalo-médico-social, assurance et prévoyance. En juillet 2012, il valide une thèse doctorale en Pathologie Humaine sur : « Le modèle d'analyse éthique des systèmes d'information en santé appliqué à la cancérologie », à l'Université de médecine d'Aix-Marseille, et financé par la société Keosys.

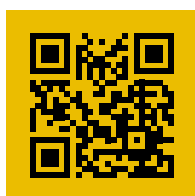
Il est actuellement Chief Scientific Officer au sein d'ADEL et chercheur associé (PhD) à l'UMR 1027 de l'Inserm. Ses recherches sont centrées sur l'approche éthique et humaine autour de la conception, la mise en place et l'usage des NTIC en santé (e-santé, m-Health, Big Data, NBIC, IA, etc.). Il a également publié une quarantaine d'articles scientifiques dans des revues reconnues, ainsi que deux livres intitulés respectivement : « Les systèmes d'information en santé et l'éthique » (ISTE, Mars 2015), et « Les Big Data et l'éthique » (ISTE, Janvier 2016). Ainsi, après avoir appliqué ses travaux de recherche pour l'usage interne de Keosys, Jérôme Béranger a cofondé le label éthique ADEL pour les systèmes d'information automatisés et les projets Big Data. Suite à une première interview à retrouver dans l'édition 2015/2016, nous avons souhaité connaître l'évolution de ce projet.



## POUVEZ-NOUS RAPPELER L'OBJECTIF DU PROJET ADEL - HEALTH ?

Le projet ADEL - Health souhaite fournir un cadre de réflexion et de bonnes pratiques éthiques sur l'élaboration, la mise en place et l'usage des données numériques de santé afin de mieux les encadrer, les contrôler et les suivre au sein des entreprises. L'objectif est de donner aux acteurs concernés par ces Big Data, les premières clés de lecture pour leur permettre d'acquiescer, d'une part, une approche éthique des NTIC, et d'autre part, de définir un cadre humain qui gère l'intérêt général et les droits individuels des données à caractère personnel. **Le label ADEL - Health permet d'acquiescer un nouvel espace de confiance des données de santé** en apportant du sens et de la sécurité tout le long du cycle de vie d'une donnée numérique en santé. Pour ce faire, le label ADEL permet à chaque entreprise ou établissement qui rentrera dans le processus d'évaluation d'obtenir :

- Un compte-rendu comprenant une cartographie détaillée sur l'état de leur situation pour le traitement de leurs données numériques, un score final éthique, et des recommandations de bonnes pratiques ;
- Un livre blanc intitulé : « Vade-mecum sur le traitement éthique des données numériques en santé » pour les accompagner dans leurs traitements de données numériques. Ce guide méthodologique est un atout non négligeable auprès des entreprises qui souhaitent passer avec succès les étapes de la certification ADEL. Ce livre blanc est téléchargeable gratuitement sur le site Internet <http://www.adel-label.com/>



Le projet ADEL constitue le **premier label sur l'éthique des systèmes d'information, des algorithmes de traitement et des données numériques**. Il pose les jalons d'une « éthique algorithmique », c'est-à-dire une éthique propre au numérique.

## UN AN APRÈS QUELLES SONT LES GRANDES AVANCÉES POUR VOTRE PROJET DE LABEL ÉTHIQUE ?

Nous avons constitué un Comité Scientifique qui comprend des experts indépendants pluridisciplinaires de haut niveau, tous animés par la même vision : **celle d'intégrer une approche éthique dans l'écosystème du numérique**. Le comité comprend deux présidents d'honneur : Cédric Villani (Médaille Fields 2010) et Gilles Babinet (Digital Champion pour la France), ainsi que 17 membres parmi lesquels des Data Scientists, professionnels de la sécurité autour des données numériques, institutionnels, mathématiciens, professionnels de la santé, juristes, philosophes, sociologues, entrepreneurs, réseaux d'e-patients et d'e-médecins. Leur mission est d'apporter des avis consultatifs et des ajustements sur la méthode et la grille d'évaluation éthique du programme scientifique ADEL afin de rendre l'outil le plus pertinent et cohérent possible.

Le projet ADEL a également obtenu de nombreux soutiens, par exemple de l'IUCT-Oncopole, l'École Polytechnique de l'Assurance (EPA), la BPI, les cabinets d'avocats Kamkar & Williate et Zorn Avocat etc.

Nous avons construit un processus complet d'évaluation comprenant les différentes grilles de questionnaires. Chaque projet impliquant le traitement de données numériques rentre dans un dispositif déclaratif de certification semi-automatisée instauré par ADEL.

- Etape 1 : Pré-questionnaire qualitatif d'entrée de 20 questions
- Etape 2 : Rapport d'entrée
- Etape 3 : Questionnaires d'évaluation éthique (sur le traitement numérique et les moyens mis en place) composés de 138 questions et de 21 questions sur la mise en place
- Etape 4 : Compte-rendu automatisé (scores, résumé, cartographies, et préconisation)
- Etape 5 : Pièces justificatives demandées
- Etape 6 : Rapport final
- Etape 7 : Obtention ou non du label ADEL - Health

Enfin l'arrivée des premiers établissements et structures qui ont éprouvé et testé le label éthique ADEL - Health, marque également une étape importante. C'est notamment le cas du centre de radiologie de Val du Cens (Saint-Herblain) et de la clinique Pasteur (Toulouse).

## QUELS SONT LES PRINCIPES ET CRITÈRES D'ÉVALUATION DE LA VALORISATION ÉTHIQUE DE LA DONNÉE ?

### Cible éthique des données de santé à caractère personnel



Le label éthique se traduit par un questionnaire détaillé et précis d'une centaine de questions autour de la valeur intrinsèque de la donnée dès la conception, de la valeur de gestion et de mise en place de la donnée et enfin de la valeur d'exploitation ou d'usage. Notre approche éthique se fonde ainsi sur les principes éthiques de Beauchamp et Childress (2001), à savoir **les principes de bienfaisance, autonomie, non-malfaisance et justice** ainsi que 8 autres domaines qui forment ce que nous nommons « **Infoethics Management** » (Cf. encadré). Chacun de ces principes et domaine fait l'objet de questions parmi lesquelles :

- Quels sont les objectifs, les buts, les enjeux et le sens de l'utilisation des données ?
- Que vais-je utiliser comme données ? Des données partielles ou totales ?
- Comment vais-je les utiliser ? A quel endroit ? Auprès de quels utilisateurs ?
- Plus globalement, comment exploiter l'ensemble hétérogène de ces données médicales accumulées et stockées dans un SI ?
- Quelle sera leur pertinence par rapport à ma situation ?
- Cela ne va-t-il pas dénaturer la valeur informative initiale ?
- L'intégrité du message final sera-t-elle conservée ?
- Cela rentre-t-il dans le cadre d'un meilleur usage du SI, d'une communication médicale optimisée et d'une amélioration de la prise en charge des soins vis-à-vis du patient ?

## Illustration des principes de l'«Infoethics Management»

### BIENFAISANCE : LE BIEN-FONDÉ D'UNE ACTION

- Est-ce que votre exploitation des Big Data en santé œuvre bien dans l'intérêt du patient ?
- L'exploitation de vos Big Data aide-t-elle à la prise de décision établie par le professionnel de santé ?

### JUSTICE : ÉQUITÉ, ÉGALITÉ DANS LE TRAITEMENT DES DONNÉES

- Vos Big Data permettent-elles un accès différent à l'information suivant la situation ?
- Est-ce que les Big Data sont disponibles en Open Data, Open Source ?

### NON MALFAISANCE ET MINIMISATION DES RISQUES

- Est-ce que l'exploitation des données respecte la protection des données ?
- L'exploitation des données s'assure-t-elle de la pertinence des outils ?

### AUTONOMIE

- L'organisation s'adapte-t-elle rapidement à l'exploitation quotidienne des données ?
- Respecte-t-elle la vie privée et le secret professionnel ?

### STRATÉGIE ET MÉTHODOLOGIE

- D'un point de vue technique, les Big Data utilisées sont-elles fonctionnelles et opérationnelles en l'état, en fonction des requêtes demandées ?
- Dans le cas d'un Système d'Information automatisé : avez-vous effectué des postulats de départ, énoncés et décrits, auxquels est confronté votre SI ?

### ORGANISATION ET RÉGLEMENTAIRE

- Les Big Data utilisées sont-elles toutes pertinentes ou exactes après traitement ?
- Un inventaire des données disponibles ou mobilisables est-il réalisé ?

### STRUCTURE ET TECHNOLOGIE

- Les Big Data sont-elles intègres (neutres, entières, non biaisées) ?
- Le code source qui compose le SI automatisé peut-il évoluer rapidement ?

### RELATION ET CULTURE

- Les Big Data améliorent-elles l'interactivité avec les acteurs impliqués dans la prise en charge du patient ?
- Est-ce que vos équipes impliquées dans la conception, la mise en place et l'usage du SI sont régulièrement formées ?

### PILOTAGE

- Avez-vous établi un SWOT autour des enjeux de vos projets Big Data ?

### MISE EN FORME

- Existe-t-il un dispositif de partage des données de santé ?
- Avez-vous bien délimité le périmètre de vos données à caractère personnel, sensible ?

### CYCLE DE DÉVELOPPEMENT

- Avez-vous bien appliqué tous les outils de sécurisation des échanges de données ?
- Est-ce qu'il y a une démarche de reporting des actions au responsable du projet ?

### OPÉRATIONS

- Existe-t-il un contrôle de la démarche qualité de sauvegarde des données de santé ?
- Les bases de données sont-elles administrées ?

## QUELLES SONT LES ENTREPRISES AYANT DÉJÀ ÉPROUVÉ LE LABEL ?

Les entreprises intéressées ou ayant testé le label sont les cabinets de radiologie (tel que celui du Val du Cens), les cliniques privées (Pasteur), les établissements publics de santé, les assurances santé, les oncopoles (Toulouse), et plus généralement des projets Big Data, Open data, etc.

Il n'y a que **20%** de contexte dans ces démarches, l'intérêt de notre démarche réside donc dans les **80%** restants qui relèvent d'une démarche duplicable à tous les projets Big Data des autres secteurs d'activité

Le label ADEL propose d'auditer différents domaines d'activité tels que les systèmes d'information et bases de données des établissements de santé, les systèmes et plates-formes informatiques, bio-statistiques et bio-informatiques gérant des bases de données et dossiers patients, l'industrie des dispositifs médicaux utilisant nécessairement des données de santé pour leur accréditation en tant que dispositif médical, les dossiers et process de gestion des essais cliniques académiques et industriels.

Il n'y a que **20%** de contexte dans ces démarches, l'intérêt de notre démarche réside donc dans les **80%** restants qui relèvent d'une démarche duplicable à tous les projets Big Data des autres secteurs d'activité comme : l'aérospatial, la finance, les transports, le commerce, la distribution, le manufacturing, les services, les utilities, les télécoms, le secteur public, l'éducation, etc.

## QUI EFFECTUE L'ÉVALUATION DES DIFFÉRENTS CRITÈRES ET LE SCORING ?

L'idée du projet ADEL est de développer un outil d'évaluation éthique qui soit totalement automatisé. **L'approche veut tendre vers une sorte d'ubérisation d'un audit éthique du traitement des données numériques. Je suis persuadé que l'intégration d'un cadre humain dans l'écosystème digital passe nécessairement par une numérisation de l'éthique.** C'est ce que je nomme l'éthique algorithmique, c'est-à-dire la conception d'un algorithme qui évalue l'éthique du cycle de vie des données numériques (de l'acquisition à la valorisation, en passant par le stockage et le traitement). Pour l'instant, le process d'évaluation du label ADEL est semi-automatisé. En effet, la vérification des livrables demandés aux structures qui éprouvent le label est effectué par un spécialiste du secteur.

## QUEL EST L'INTÉRÊT POUR LES ENTREPRISES À SE SOUMETTRE À VOTRE LABEL ÉTHIQUE ?

En ce qui concerne les acteurs de la e-santé, de la médecine 3.0 et 4.0, les établissements de santé, les industries pharmaceutiques, le label ADEL - Health a pour principal avantage de les accompagner dans l'exploitation et l'usage de leurs données numériques de santé, en apportant un guide de bonnes pratiques et un cadre pragmatique et humain. En ce qui concerne les assurances et mutuelles santé, banques, opérateurs, etc., le label ADEL a pour intérêt d'apporter de la confiance et une garantie de bon usage des données numériques auprès de leurs clients et des citoyens.

**LE BUT FINAL DE VOTRE LABEL EST DE PERMETTRE UNE GESTION ET GOUVERNANCE RÉUSSIE DE DONNÉES. SI VOUS DEVIEZ CITER LES RECOMMANDATIONS PRIORITAIRES POUR Y PARVENIR, QUELLES SERAIENT-ELLES ?**

Plusieurs mesures et recommandations s'imposent à nous afin de garantir la qualité et la protection des données de santé à caractère personnel. Elles sont résumées dans l'encadré ci-dessous.

**Recommandations entourant la mise en place et l'exploitation de la donnée de santé à caractère personnel**

- Indiquer l'origine de la donnée : brute, consolidée via différents producteurs, construites à partir de tierces données ;
- La donnée de santé à caractère personnel ne peut être collectée et traitée que pour un usage déterminé et légitime ;
- Seules les données pertinentes et nécessaires aux objectifs d'exploitation doivent être traitées ;
- La donnée médicale doit circuler en garantissant un souci de confidentialité ;
- Indiquer clairement la date de transfert ou d'hébergement des données ;
- Chiffrer la donnée pour assurer une traçabilité et un suivi de cette dernière ;
- Les outils de transmission de l'information doivent être valides ;
- La donnée médicale est sans cesse évolutive et doit être régulièrement mise à jour ;
- Fournir de l'information consistante et basée sur des preuves médicales ;
- Indiquer clairement si l'information est fondée sur des études scientifiques, consensus d'experts, ou sur une expérience ou une opinion professionnelle ou personnelle ;
- S'assurer que les avis médicaux ou conseils sont donnés par des praticiens qualifiés ;
- Décrire le processus de recrutement : typologie des contributeurs, contrat avec le fournisseur, mode de rémunération, consentement de la personne ;
- Indiquer la modalité de fabrication, de constitution ou de transformation de la donnée ;
- Décrire le processus d'usage et de conservation de ses données médicales par le fournisseur ainsi que les conditions dans lesquelles l'utilisateur pourra lui-même accéder à ses data ;
- Les données personnelles de santé ne doivent pas être conservées indéfiniment et leur durée de conservation est appréciée en fonction de la finalité du traitement ;
- Évaluer rigoureusement et équitablement l'information ;
- Respecter les droits associés à la protection intellectuelle ;
- S'assurer que les fournisseurs respectent les réglementations génériques et particulières liées à l'exploitation de la donnée personnelle ;
- Une cohérence de la donnée nécessite un temps de coordination ;
- Les produits ou services employés doivent être décrits dans un langage clair, facile à lire et approprié aux utilisateurs ciblés.

**LA RÉGLEMENTATION EUROPÉENNE SUR LES DONNÉES PERSONNELLES DU 14 AVRIL 2016, VA-T-ELLE SELON VOUS AVOIR UNE INFLUENCE SUR UNE UTILISATION PLUS ÉTHIQUE DE LA DONNÉE ?**

Je ne suis pas sûr que la récente réglementation européenne en vigueur va avoir une influence importante sur un usage plus éthique de la donnée. Il faudrait, selon moi, une loi imposant à toutes les structures et organisations, qui traitent de la donnée numérique, **l'instauration en leur sein d'un Comité opérationnel d'éthique pluridisciplinaire pour piloter la conception, la mise en place et l'usage des Big Data en santé.**

**VOTRE LABEL ÉTHIQUE EST BASÉ SUR UNE PARTICIPATION VOLONTAIRE DES ENTREPRISES, PENSEZ-VOUS QU'UNE RÉGLEMENTATION PUISSE RENDRE UN JOUR UN LABEL ÉTHIQUE OBLIGATOIRE ?**

Je suis de nature optimiste ! Comme le juridique s'alimente et s'inspire des réflexions éthiques pour acter et légiférer, j'ai bon espoir ... De plus, je remarque que ces dernières années, les séminaires et congrès autour du numérique donnent une place de plus en plus importante à l'approche éthique autour du digital. En effet, après une phase où les acteurs ont dû digérer la compréhension et la maîtrise des NTIC et des data sciences, nous rentrons dans une ère où ces derniers prennent un peu plus de hauteur pour établir des réflexions sur les impacts, les enjeux, les risques au niveau sociétal.

Dès lors, **une réglementation imposant un label éthique obligatoire serait une avancée considérable.** Dans ces conditions, une prise de conscience de la valeur des données issues des NTIC devient essentielle, tout comme **un meilleur encadrement de l'usage des algorithmes de traitement à l'égard des individus.** En effet, si nous voulons vivre harmonieusement dans ce nouveau monde régi par la « gouvernance algorithmique », nous devons intensifier nos efforts : **sur l'utilisation des données, qui permettent au Machine Learning d'apprendre, sur la maîtrise des données par l'utilisateur et sur la transparence de leur utilisation.**

L'«éthique algorithmique» devient décisive afin que les professionnels des Big Data trouvent un juste équilibre entre le traitement rationnel et mesuré des données, dans le respect de la vie privée et un surtraitement qui serait contre-productif, notamment dans la relation médecin-patient.

**L'étape ultime serait d'avoir un algorithme qui parle à un algorithme, une uberisation de l'audit et de la certification avec une vérification semi-automatisée, voire totalement automatisée, par un algorithme de traitement afin que le numérique combatte lui-même ses propres failles.**



**CÉCILE MONTEIL**

Médecin aux Urgences Pédiatriques  
Fondatrice de l'association Eppocrate  
Bloggeuse santé et animatrice du  
**Zapp' du Doc / Quotidien du Médecin**  
**HÔPITAL ROBERT DEBRÉ**  
**PARIS**

«A l'ère du Big Data, de l'Intelligence Artificielle et de la robotique, c'est une médecine ultra-technologique de performance et de précision qui s'offre à nous. Mieux soigner, pour vivre mieux et plus longtemps. Mais pourquoi s'arrêter à la réparation de l'Homme quand on peut aussi l'augmenter ?

La discipline médicale entre alors dans une nouvelle dimension sociétale : quelle Humanité allons-nous façonner pour demain ?

Le Big Data sera-t-il utilisé comme un outil de surveillance ultime pour une société de normopathes bien portants ? Est-ce que l'on hybridera corps et esprits avec la machine, ou bien nous contenterons-nous de sélectionner et reprogrammer génétiquement l'humain ?

Le danger n'est pas dans la technologie, mais bien chez l'Homme qui en a le contrôle. La réflexion éthique est indispensable, et le projet ADEL est essentiel à ce titre ».



### GILLES BABINET

Digital Champion France /  
Président d'honneur

#### COMMISSION EUROPÉENNE / COMITÉ SCIENTIFIQUE D'ADEL

« Mettre l'éthique au cœur de la modernité ; s'en servir comme d'un axe pour permettre à la révolution digitale de se déployer dans l'univers de la santé me semble être une approche pertinente autant qu'utile.

Toute la difficulté consiste de ne pas adopter de posture manichéenne : trop de conservatisme nous empêchera de rentrer dans cette révolution tandis qu'une logique d'innovation débridée pourrait nous faire perdre une partie de ce qui nous caractérise en tant qu'humanité.

L'éthique, qui privilégie une vision d'ensemble, et non pas le point de vue du juriste contre celle de l'innovateur me semble répondre à ces défis.

C'est pourquoi j'apprécie et soutiens l'approche et les travaux de Jérôme Béranger dans le projet ADEL ».



### LINA WILLIATTE PELLITTERI

Professeur FLD-UCL  
Membre du bureau exécutif de la  
Société Française de Télémedecine  
(SFT ANTEL)

#### AVOCAT BARREAU DE LILLE CABINET WT AVOCATS

«Les objets connectés font désormais partie de notre quotidien. Qu'il s'agisse de la montre qui révèle à la fin de la journée la fréquence normale ou non de notre rythme cardiaque, ou du téléphone portable qui, en indiquant le nombre d'appels manqués, nous informe du nombre de pas réalisés lors la journée. Si leur intérêt est incontestable, la traçabilité des données qu'ils génèrent questionne le juriste. Nul doute, que ces informations récoltées, traitées et classées dans la catégorie des données de « bien être » desscellent incidemment des informations précieuses sur notre santé.

Or les données de santé sont qualifiées en France de données sensibles, dont le traitement est interdit par la loi informatique et liberté de 1978 modifiée récemment en 2016, sauf avec le consentement de la personne.

Ce consentement, pour être valable, doit être donné en connaissance de

cause. C'est ce point qui interpelle, car si la personne accepte d'être tracée sur sa forme physique, il est à craindre qu'elle refuse de l'être sur ses données de santé.

C'est pourquoi, l'objet connecté non qualifié de dispositif médical mais simplement d'objet permettant le quantified self peut être une menace pour la garantie des droits des personnes sur la confidentialité de leurs données de santé à caractère personnel.

Cette menace très souvent ignorée par les utilisateurs finaux peut l'être également par les industriels, qui financent leur mise sur le marché. C'est pourquoi, il faut saluer et encourager des initiatives telles que celles prises par les auteurs du Label Adel Health qui en permettant d'analyser toutes les données récoltées via les objets connectés, donne la possibilité aux industriels de se rendre compte des détournements potentiels de finalité que l'utilisateur pourrait subir et les aide ainsi à mieux connaître leurs produits avant la mise sur le marché».

BENOÎT  
**BINACHON**

Managing Director

Executive Search - Smart / Big Data Analytics

**UMAN PARTNERS**



Du  
**recrutement**  
dans la **Data**  
à la Data  
dans les  
**processus** de  
recrutement





Ingénieur dans le domaine spatial de formation, **Benoît Binachon** s'est très rapidement redirigé vers le domaine du recrutement en intégrant le Groupe Michael Page avant sa forte croissance. Une aventure exceptionnelle qui lui a permis, en 2000, d'être propulsé parmi les directeurs généraux de Michael Page France. Il a pris le goût au recrutement et est resté dans le groupe, de culture entrepreneuriale très forte, jusqu'en 2004.

Il cofonde et dirige ensuite la start-up Business effiScience (40 personnes, 10M€ de CA, cédée à BearingPoint). Il s'est ainsi forgé, de 2004 à 2012, une expérience opérationnelle rare et reconnue dans le domaine de la Data Science, du Big Data et du Machine Learning en développant notamment de nombreux cas d'usage innovants et une boîte à outils d'algorithmes très pointus.

Il a enfin souhaité revenir dans le domaine du recrutement par passion et a ainsi cofondé Uman Partners, après un an et demi passé au sein de Korn Ferry.

### **QU'EST-CE QUI VOUS A POUSSÉ À FONDER UMAN PARTNERS ?**

Nous avons fondé **Uman Partners** sur deux dimensions importantes, tout d'abord le **recrutement dans la Data**, par goût et car il s'agit d'une spécialité métier intéressante et d'un véritable besoin des entreprises, mais également la volonté d'insérer de **la Data et du quantitatif dans nos processus de travail**. Notre approche du discernement des candidats est donc très « data driven ». Notre objectif au sein de Uman Partners est d'aider nos clients à recruter dans tous les métiers seniors liés au Big Data, à la Data Science et au Machine Learning. Nous ne recrutons pas de profils quantitatifs classiques mais des Data Scientists Seniors, des patrons de la Data Science, des Chief Data Scientists...

### **SUR QUELS TYPES DE PROFILS VOUS SOLLICITE-T-ON LE PLUS ?**

Nous recrutons trois types de profils, les premiers et plus recherchés sont liés à **la Data Science pure et dure**, les métiers cœurs scientifiques ainsi que les métiers techniques autour du Big Data et de l'architecture qui souvent se mêlent aux premiers. Il y a des gens qui sont de dominante Data Science avec une mineure Big Data et des gens qui sont de dominante Big Data avec une mineure Data Science. **Le premier besoin est en effet de construire la technique et la science**, les entreprises recrutent d'abord le Chief Data Scientist et ensuite les personnes qui vont orchestrer le projet.

Les seconds profils recherchés sont les utilisateurs et intermédiaires de la Data science, soit des consultants qui savent parler le double langage Data science et métier : des directeurs de projet qui vont conduire des initiatives Data pour le compte du marketing, du risque, du consumer insights...

Il s'agit de personnes à même de comprendre le besoin d'un directeur marketing sur sa pro-

blématique de fidélité, de développement d'un nouveau produit, ou d'un directeur des fraudes sur une problématique de fraude à l'assurance par téléphone.

On constate, de la part des industriels et entreprises en général, une minimisation de l'importance de recruter des chefs de projets experts de la Data. Les entreprises ne sont pas conscientes de l'importance de la sensibilité des directeurs de projets et patrons de conseils à la Data. Mais il est vrai qu'il est possible de façonner un chef de projet Data Science à condition qu'il y soit sensible à la base. Une personne n'ayant pas forcément eu un parcours lié à la Data Science, sera capable d'assurer ce poste si elle est en mesure de détecter les cas d'usage qui peuvent être résolus par la Data Science à partir des outils disponibles. On trouve des personnes diplômées de Science Po qui vont en être capables alors que des purs mathématiciens vont y être insensibles.

Les derniers postes pour lesquels nous sommes sollicités sont ceux qui demandent une forte sensibilité à la Data : Directeur Marketing Client, Directeurs Consumer Insight etc. Aujourd'hui, on demande par exemple au Directeur Marketing de gérer un call center avec une approche très quantitative. On peut également demander à un patron du Consumer Insight dans la cosmétique ou en pharmacie de savoir introduire de la prédiction, de la prescription dans l'élaboration d'un parfum, d'une crème de beauté car cela coûte très cher de faire des essais et de développer des produits sans avoir une visibilité du succès d'un produit sur une population. La Data donne une grande longueur d'avance dans les fonctions métiers. On ne demande pas à ces métiers d'être des Data Scientists mais d'avoir une compréhension de ce que l'on peut faire avec quels moyens, d'être à l'écoute et d'utiliser tout l'écosystème technique. Sur cette dernière famille de postes, on constate un vrai retard en France notamment par rapport aux Etats-Unis, qui ont

vu apparaître plus de postes data-driven. Il y a naturellement moins de besoins mais également moins de candidats sur ces postes en France et donc une grande difficulté à trouver des candidats qui ont vraiment une expérience data-driven sur leurs sujets.

### **QUELLE INTÉGRATION DU BIG DATA DANS LES FONCTIONS MANAGEMENT DE L'ENTREPRISE ?**

En France, on constate une avancée plutôt par le digital au niveau des directions. C'est le digital qui rentre au Comex. Ce n'est pas forcément réducteur car **la Data est un des instruments du digital**. En revanche, il y a beaucoup d'initiatives digitales qui ne tiennent pas compte de la Data, ce qui est complètement aberrant. **Faire du digital sans faire de la Data, c'est comme construire une voiture sans moteur** et il est assez courant de voir un patron du digital ou Chief Digital Officer qui n'est pas du tout sensible à la Data. De même, **il y a encore très peu de Chief Data Officers en France** contrairement aux Etats-Unis. Le rôle du Chief Data Officer est de bien comprendre tous les métiers du comex : celui du Directeur Marketing, Financier, de la R&D, de la Production etc. Il doit être une éponge pour comprendre tout l'écosystème de la Data, les besoins de son entreprise et l'ensemble de l'offre qui existent sur les sujets Data pour servir ces besoins.

Malgré son rôle primordial, l'important n'est pas qu'un Chief Data Officer porte le projet impérativement, mais bien que **la Data soit une des priorités du Comex de l'entreprise**.

Faire du **digital sans faire de la Data**, c'est comme construire **une voiture sans moteur**

### **FACE À DES GÉANTS COMME FACEBOOK OU GOOGLE, COMMENT VOS CLIENTS ARRIVENT-ILS À RESTER ATTRACTIFS POUR DES PROFILS DATA SCIENCE ET BIG DATA QUI SE RARÉFIENT ?**

Google ou Facebook vont attirer des profils très scientifiques, spécialisés par exemple dans la classification ou l'indexation de vidéos. Il s'agit finalement de postes avec un périmètre assez restreint, qui conviendront à des purs geeks.

Lorsque nous accompagnons nos clients sur le lancement d'un Data Lab par exemple, il s'agit d'un véritable projet de transformation de l'entreprise. Pour les banques, opérateurs téléphoniques, assureurs, il s'agit de construire une équipe avec un périmètre très large. Les enjeux scientifiques sont tout aussi pointus mais il y a également des enjeux humains, des notions de transformation et du management.

Lorsqu'Axa lance son Data Innovation Lab, il s'agit de transformer toute la chaîne, de la captation du sinistre, des données du sinistre jusqu'à une action du téléopérateur, du

conseiller client dans un call center. De même, pour un opérateur télécom, un des cas d'usage est de faire apparaître sur l'écran du téléopérateur une instruction très précise qui prescrit une action à appliquer au client qui appelle. Cela implique de la Data Science en amont, du juridique sur les données exploitables, de l'informatique, du syndical, car on fait évoluer la manière de travailler, et de la formation car cela a un impact sur l'action des téléconseillers qui vont obéir aux ordres de l'algorithme. Ces deux derniers cas montrent que les grands groupes restent attractifs car il s'agit d'une conduite du changement très large et dans laquelle la dimension humaine est très intéressante pour un Data Scientist.

De plus, les grandes entreprises telles Facebook, Google... sont américaines ou étrangères et les missions sont plus limitées pour un Data Scientist qui travaillera pour eux en France.

**En revanche, un vrai handicap pour les entreprises françaises réside dans la pression salariale exercée par Google, Facebook ou encore GE qui offrent des salaires bien supérieurs aux salaires consentis aux Data-Scientists dans la majorité des entreprises françaises. De même et grâce à l'attractivité des salaires proposés, les entreprises américaines attirent de plus en plus de Data-Scientists français, dont les compétences techniques sont au plus haut niveau mondial. Pour éviter un brain drain (fuite des cerveaux) de notre vivier en Data-Science et continuer à attirer des talents, les entreprises françaises doivent nécessairement revoir leurs barèmes de salaires.**

### **QUELLES NOUVELLES COMPÉTENCES SONT DÉSORMAIS DEMANDÉES POUR UN SPÉCIALISTE BIG DATA ET DATA SCIENCE ?**

Dans cet univers, ce sont des profils qui seront toujours rares... Même si on forme beaucoup plus de profils à la Data au sein de Polytechnique, Supélec, HEC... Ce sont des métiers élitistes qui produisent peu de talents car ils demandent beaucoup de capacités d'abstraction et de créativité, des qualités qui sont innées et sélectives.

Les compétences requises dans ce domaine sont de deux natures. On demande de nouvelles compétences techniques en lien avec de plus ou moins nouveaux langages et formalismes mathématiques. La Data Science est avant tout composée de gens qui sont très mathématiques, qui comprennent les concepts mais qui sont également capables d'avoir un véritable recul scientifique. Faire de l'informatique sans garde-fou scientifique peut s'avérer dangereux.

Mais l'on demande désormais également aux talents une vraie capacité à comprendre le business et une capacité à participer à la transformation. Ce sont des compétences qui relèvent du leadership et qui incluent en fonction des postes et niveaux de :

- Savoir influencer un dirigeant ;
- Savoir porter l'innovation ou détecter les innovations qui peuvent être appliquées à une micro ou macro échelle ;
- Porter et développer ses équipes – savoir les manager avec les méthodes inspirées par les Google ou autres – savoir les faire rayonner, les connecter à la recherche, les faire parler dans des meetups, en plus de les faire rayonner au quotidien en les aidant à se développer d'un point de vue strictement scientifique ;
- Savoir vite incorporer dans leur disque dur cérébral de nouvelles approches et nouvelles technologies ;
- Avoir le sens du client et parler avec des mots simples de concepts abstraits qui ont des applications très concrètes dans les métiers ;
- L'écoute...

Les talents dans le domaine de la Data Science sont des moutons à cinq pattes. On peut imaginer le super Data Scientist comme introverti et peu communicant, alors qu'il s'agit d'une compétence primordiale. **La Data Science n'a de sens qu'en face d'un métier à qui il faut savoir parler.**

Pour exemple, dans la fraude, le faux positif (le fait de détecter quelqu'un comme fraudeur

alors qu'il ne l'est pas) est très préjudiciable et peut avoir un impact business très négatif pour une entreprise. L'enjeu pour un Data Scientist est donc d'en minimiser le nombre et d'en comprendre les enjeux. Une fraude à la carte bleue (dont le client est victime) diffère d'une fraude sur un compteur électrique, dans laquelle l'accusation de fraude est faite sur le client lui-même.

Pour les métiers intermédiaires, la compétence primordiale réside dans la culture de la Data Science (et du Big Data) qui va leur être demandée d'emblée.

Enfin pour les métiers utilisateurs de la Data comme le Marketing, une des compétences les plus importantes sera de savoir porter l'innovation, détecter l'innovation, sa valeur marchande, la projeter et adapter la technologie dans leur métier à priori non technique.

### **LES GOUROUS DU BIG DATA PARLENT DE LA MORT DES FAUX DATA SCIENTISTS ET DE LA NAISSANCE DES DATA SCIENTISTS CONVERTIS. EST-CE QUE VOUS OBSERVEZ CES PHÉNOMÈNES ?**

Selon moi, il n'y a pas de faux Data Scientists. Il y a les disciplines des statistiques, de l'actuariat, de l'économétrie ... et celles de la Data Science et du Big Data. Il s'agit de deux mondes assez distincts (mais qui ont des zones communes) qui répondent à des besoins différents. Et il y a simplement des gens moins appétants au Monde de la Data Science parmi les statisticiens, économètres, actuaires ou même les profils en mathématiques financières, et d'autres qui s'y convertissent. Et d'autres encore qui ont toujours été dans la data science (qu'on appelait avant Intelligence Artificielle née dans les années 70). Je pense que ce que l'on pourrait appeler les faux Data Scientists (quoique très péjoratif de les qualifier de « faux ») seraient des personnes qui sont d'excellents utilisateurs de boîtes à outils existantes, champions de challenges Kaggle, mais qui n'ont pas forcément le recul suffisant au niveau scientifique. Ils sont cependant indispensables à l'écosystème de la Data Science. Le Data Scientist, l'informaticien et l'Architecte Big Data sont tous complémentaires.

Il existe bien en revanche des Data Scientists convertis, et de nombreuses voies peuvent mener à la Data Science avec comme tronc commun ce-



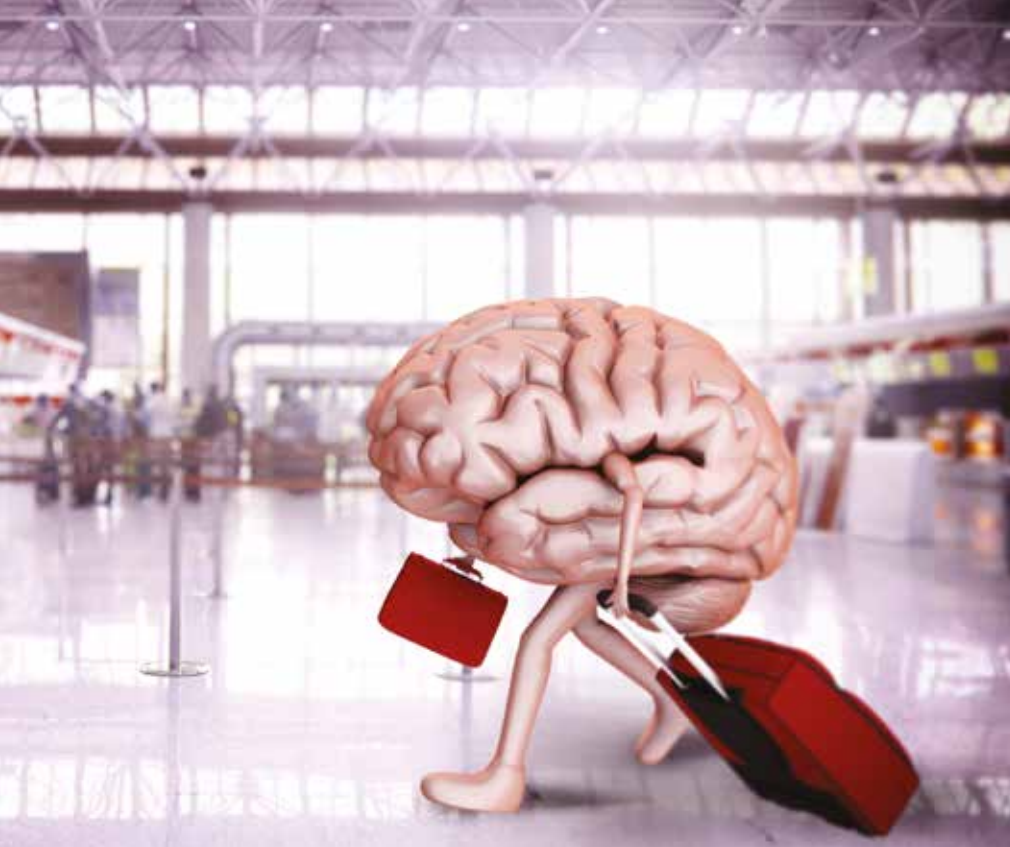
pendant qu'il s'agit toujours de scientifiques ayant une forte capacité d'abstraction, parmi lesquels :

- **Des professeurs de mathématiques** qui sont maintenant conseillers de Data Lab ;
- **Des spécialistes du traitement du signal (radar, imagerie, télécoms) ;**
- **De jeunes physiciens** qui ont pu faire récemment des thèses en physique des particules, en astrophysique et qui ont traité de tels volumes de données qu'ils ont utilisé de la Data Science et du Big Data ;
- **Des spécialistes de la physique** des particules ou de la finance quantitative haute fréquence, qui sont restés très mathématiques. Les premiers Data Scientists étaient en effet des traders ;
- **Des informaticiens** qui rentrent dans le domaine par les algorithmes.

### **VOUS UTILISEZ VOUS-MÊME UN OUTIL PRÉDICTIF DE LA RÉUSSITE DES CANDIDATS (EBI - EVIDENCE BASED INTERVIEWER), QUEL EN EST LE FONCTIONNEMENT ?**

L'esprit humain a tendance à provoquer le clonage. D'abord pour des raisons psychologiques, mais également par manque de capacités pour croiser et trier pertinemment les données réellement prédictives de la réussite. Dans le secteur des Ressources Humaines, on recrute majoritairement sur la base d'intuitions. Et contrairement aux idées reçues, la Data Science appliquée aux RH, ouvre des opportunités.

Depuis plus de 40 ans, les Nord-Américains en particulier ont conçu des outils de prédiction de la réussite d'un individu dans son futur poste en fonction de ses compétences de leadership, des caractéristiques de l'entreprise et du poste. Les algorithmes permettent de déterminer des scénarios de réussite d'un candidat. Dans le volume et la masse, les algorithmes vont mettre à jour des milliers de configurations différentes. Ils vont détecter des niches de personnes qui pour des raisons communes sont en réussite, raisons qui n'auraient pas forcément pu être relevées par l'œil humain. La détection de signaux faibles de ces niches de personnes en



La **Data Science** c'est un outil pour faire fructifier le **hasard**

réussite va permettre d'agréger toutes les personnes qui présentent les mêmes caractéristiques et de les mettre en réussite également. Si l'on détecte 15 personnes qui sont en réussite pour la même combinatoire de raisons, dont par exemple une formation, on va pouvoir appliquer le facteur de réussite à 200 000 profils jumeaux. On fabrique ici bien des clones mais pour créer des parcours qui n'auraient été générés que par le hasard : **La Data Science c'est un outil pour faire fructifier le hasard.**

En revanche, une bonne approche analytique au service des RH doit éliminer les clones de bases. On sait que salaires élevés et diplômes sont souvent corrélés. Les algorithmes suppriment donc cette corrélation pour s'intéresser aux corrélations complexes, par exemple, pourquoi certains ne réussissent pas. On détecte les signaux faibles de la réussite en enlevant les couches supérieures.

Cependant, l'humain reste primordial dans les Ressources Humaines, des recruteurs créatifs qui se fient à leurs intuitions seront toujours indispensables, car la Data Science ne détectera jamais quelque chose qui n'est jamais arrivé. Elle va détecter des signaux faibles mais pas des signaux qui n'existent pas.

Nous avons donc créé (à partir de concepts existants) une approche reposant sur l'identification de compétences dites « de leadership » aux US ou « compétences comportementales » en France) prédictives de la réussite dans un poste. Nous avons reconstruit un référentiel de 46 compétences (curiosité, influence, autonomie d'apprentissage etc...), que nous avons corrélées avec une base de données d'individus contenant les informations de parcours, les compétences de leadership et la réussite dans les postes de chacun d'entre eux.

C'est un outil qui éloigne de l'informel, qui structure l'entretien en face à face et valide les compétences de leadership pour la réussite dans le poste. Les candidats se subliment avec cette méthode en permettant au recruteur de poser les bonnes questions. Rarement un scientifique décrira la manière dont il a convaincu un dirigeant si la question ne lui est

pas posée. Il s'agit donc d'une approche plus factuelle, plus humanisante qui s'intéresse aux qualités véritablement importantes pour un poste. Cela sert la démarche d'« À compétences égales », association dont nous sommes membres. À compétences égales, nécessaires pour un poste, nous donnons toutes leurs chances aux personnes qui peuvent être introverties par exemple.

Il s'agit d'un outil data-driven supplémentaire, qui ne fait pas pour autant de notre cabinet une machine à recruter, mais au contraire permet au recruteur d'y voir plus clair sur les aspirations et capacités des candidats.





CHRISTOPHE  
**MONTAGNON**

DOSI

**RANDSTAD**



# RH & **BIG DATA :** Des outils au service de la **fluidification** du **marché** de **l'emploi**

**Christophe Montagnon** a rejoint Randstad il y a plus de 20 ans. Il a rejoint le secteur des ressources humaines (RH) comme directeur d'agences avant de poursuivre un parcours opérationnel en élargissant ses responsabilités géographiques. En 2003, il commence à travailler au sein des réseaux Randstad de spécialité, appelés centres experts (CXP). Lorsque Randstad fait l'acquisition de Védior, il devient directeur de l'Organisation France pour diffuser au sein du nouvel ensemble les principes organisationnels de Randstad. Il y a 5 ans, François Béharel, président du groupe Randstad France, lui confie par ailleurs l'organisation des Systèmes d'Information de l'entreprise.

## QUELLES CONVICTIONS ET QUELLE ÉVOLUTION DU MÉTIER DE RH ?

Nous avons une conviction extrêmement forte, que résume notre baseline « Tech & Touch » : **l'alliance de la technologie et de l'humain**. Nous sommes convaincus que notre métier va intégrer la technologie à la fois au travers des interfaces digitales, des plateformes numériques, mais aussi de la data pour fonder les décisions. La technologie ne va cependant pas faire disparaître la relation humaine, c'est pour cela que nous maintenons ces deux dimensions, avec chacune le même poids. Nous pensons que la technologie va venir s'intégrer à la fois dans la manière dont nous produisons nos services et dont nous interagissons avec les gens. Nous aurons toujours besoin d'interactions humains-humains. Nous voyons l'évolution de notre métier qui bouge vers ce Tech & Touch.

## QUEL ENGAGEMENT DE LA PRÉSIDENTE ?

Au sein de Randstad, l'engagement de la présidente sur ce sujet est entier. C'est vrai du digital de manière générale, mais **la data est un engagement particulier de notre président**. François Béharel a en effet la très forte conviction que la data changera notre métier. Pour ne pas subir cette conviction et en être un acteur, il s'y investit et nous donne les moyens pour être actif sur ce segment. Notre métier n'est initialement pas d'une grande transparence : quand vous êtes un chercheur d'emploi, vous n'avez pas forcément une vue claire de ce qui se passe dans le métier, et quand vous êtes une entreprise et que vous recherchez un candidat, vous ne disposez pas non plus d'une vue claire sur les compétences et le succès que vous allez avoir dans le sourcing d'un candidat. Le fait de collecter des informations et de les rendre tangibles par des données concrètes, que nous avons collectées ou produites, aide énormément,

même si ce n'est qu'une approche de la vérité. **L'un des éléments d'engagement de François Béharel est de rendre les décisions mieux éclairées, plus transparentes et plus réfléchies.**

## QUELS SONT LES PROJETS PRINCIPAUX DANS LESQUELS VOUS AVEZ SOUHAITÉ INTRODUIRE DU BIG DATA ?

Quand je suis arrivé dans le scope Big Data, les enjeux étaient clairs et essentiellement métiers. Je n'étais bien sûr pas seul à construire le projet Big Data, le socle opérationnel et métier était très fort.

**L'appariement entre les besoins de compétences et les besoins d'emploi est un des premiers domaines dans lequel nous avons souhaité inclure du Big Data.** Dans chacun des territoires, nous regardons comment se situent l'offre et la demande pour avoir une idée plus précise et transparente des ressources disponibles, des besoins des différentes entreprises. Nous souhaitons permettre aux acteurs qui recherchent des compétences de savoir si elles sont présentes sur le territoire, d'avoir une meilleure idée de comment les sourcer, s'il y a des entreprises en recherche des mêmes profils, des entreprises en compétition, etc. Si l'on se place dans la position d'une personne en recherche d'emploi, il s'agit de leur fournir l'information sur les entreprises en recherche des compétences dont elle dispose. Ce besoin d'appariement correspond à une demande récurrente, car notre métier consiste à produire pour les entreprises de la recherche de compétences, et pour les gens qui recherchent un emploi une passerelle vers les métiers. Nous mesurons l'offre et la demande dans chacun des territoires pour cartographier les besoins.

**Nous avons intégré des processus Big Data dans le domaine des mobilités de métier à métier,**

nous avons regardé comment nous pouvions faciliter un mouvement d'un métier vers un autre métier, par ce que nous appelons des passerelles de compétences. Nous avons notamment travaillé sur des data visualisations, pour rendre graphique et simple cette logique de proximité de métiers à métiers au travers d'une taxonomie de compétences, récupérée à l'intérieur du Big Data. Nous avons donc commencé par le projet Big Data autour des appariements entre la compétence et les besoins des entreprises et en second nous avons développé les algorithmes de matching. Ce sont les deux sujets clefs sur lesquels nous avons le plus d'avance car ils produisent du service au client.

## QUID DE L'INTERNE ?

Lorsque vous décidez de **devenir une entreprise « Data-driven »**, ce qui signifie prendre des décisions fondées sur des données, des chiffres, il ne s'agit pas uniquement d'appliquer le Big Data dans le champ du service rendu au client, **il faut également l'introduire dans le champ interne** et notamment dans le CRM pour les décisions de segmentation, de pricing et les algorithmes de matching pour aider les candidats à cibler les bonnes entreprises et les bons postes. Dès lors que vous décidez que c'est en étant data-driven que vous serez plus efficient, il n'y a aucun segment de l'entreprise qui y échappe. Nous avons fait pas mal de travaux d'analytique interne, les sujets CRM sont relativement bien couverts. La donnée ne réside pas que dans les services mais aussi dans l'efficacité interne.



**QUELS SONT LES CAS D'USAGE ET LES TYPES D'ACTEURS CONCERNÉS ?**

Nous utilisons le Big Data pour répondre à des questionnements RH. Nous adressons, tout d'abord, **les chercheurs d'emploi**, pour répondre à la question suivante : est-ce que dans tel bassin, avec mon métier, j'ai une chance de trouver un emploi ? Mais cette question peut être scénarisée, par exemple si j'accepte de faire plus de déplacements dans la journée ou si j'acquière des compétences complémentaires pour évoluer d'un métier A vers un métier D. Si tel est le cas, est-ce que cela change mes opportunités ?

D'autre part, nous fournissons **aux entreprises** des réponses à des questions telles que : si je m'installe dans tel bassin d'emploi et que je recherche tel type de compétences, est-ce que j'ai une chance de les trouver ? Si oui avec qui suis-je en compétition sur cette compétence ? Est-ce que je suis représenté, visible sur ce marché ? Si je source ailleurs, est-ce que j'augmente mes chances ? Quels leviers je peux actionner ? Nous leur fournissons des réponses

sur les facteurs géographiques, les types de compétences, les types d'emploi et la compétition sur le marché de leur recherche.

Enfin, nous avons des cas d'usage **institutionnels**, nous pouvons par exemple orienter une politique de formation vers l'emploi, en fournissant des données qui permettent d'être certain de cibler les bonnes populations et les bonnes compétences à acquérir par celles-ci. Nous travaillons également avec les CCI, les Agences de développement économique et les acteurs qui s'intéressent au remplacement de personnes en chômage de longue durée, comme les maisons de l'emploi par exemple.

Dans le cadre du Big Data, nous allons interroger la plateforme sur ces questions et essayer de créer des scénarios types, sur lesquels nous n'avons pas de réponse immédiate jusqu'ici.

La temporalité est une notion complémentaire intégrée sur la plateforme. Nous avons de la donnée stockée depuis longtemps, qui nous permet de déduire des tendances et de les projeter pour faire ressortir des saisonnalités,

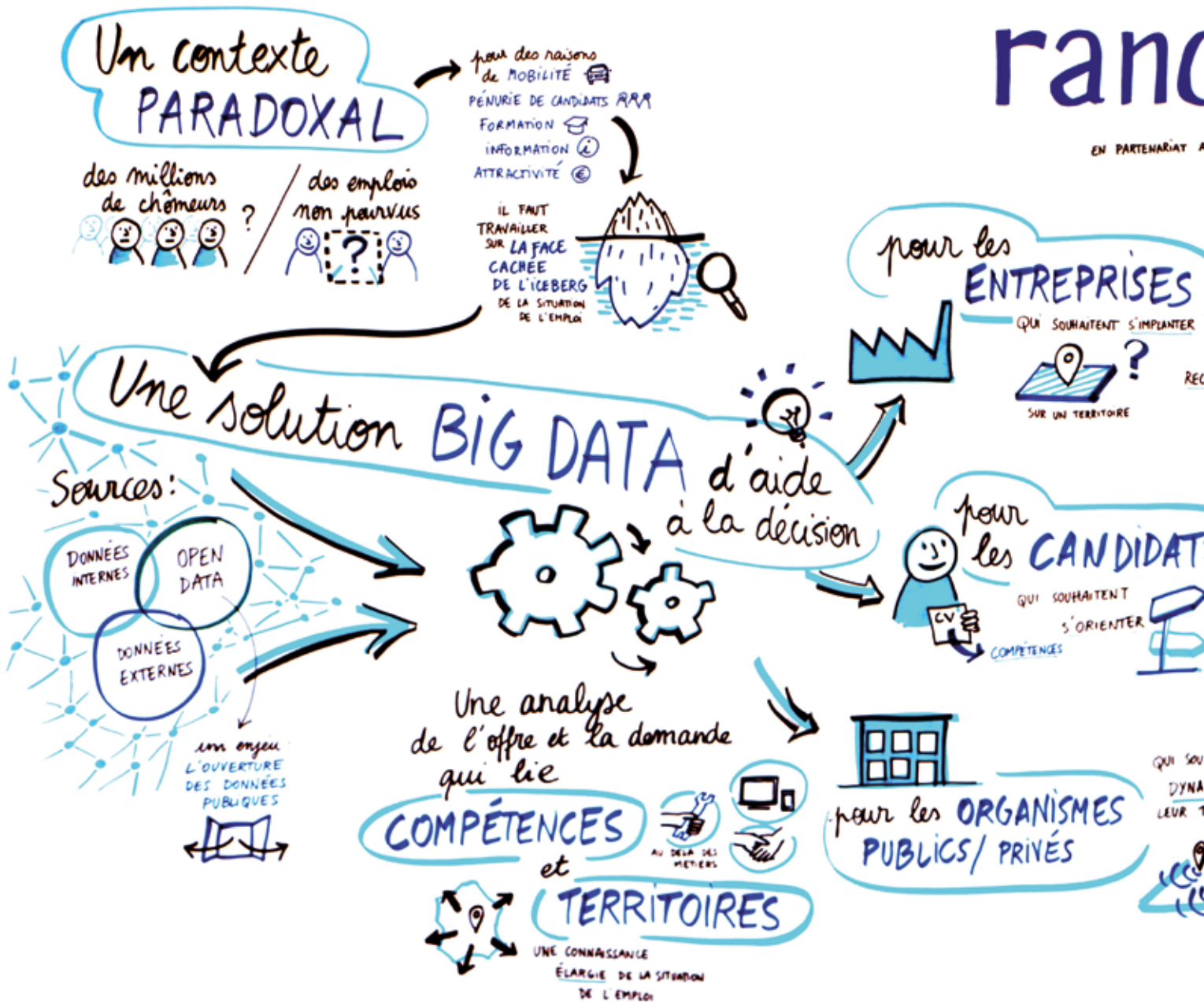
chercher des récurrences, des corrélations. Le volume d'informations stockées nous permet d'imaginer, non pas des cas d'usage et des questions à un instant T, mais des cas d'usage dans un temps projeté et de faire des prédictions.

**QUELS TYPES DE DONNÉES RÉCOLTEZ VOUS ?**

Une grande partie de nos data sont des données structurées internes, mais nous disposons également de données non structurées sémantiques, liées à tout ce qui est récolté via notre portail ainsi que via les CVs. Nous récupérons des données externes d'open source ou d'open data. Nous achetons par ailleurs des données essentiellement de Crowling et obtenons des données gratuites officielles, de type INSEE et Pôle Emploi. Nous n'utilisons pas encore les données de réseaux sociaux, mais sommes en réflexion pour les intégrer.

**QUELLE EST L'INFRASTRUCTURE CHOISIE ?**

Nous avons une infrastructure de recette basée sur du Hadoop traditionnel en clusters et une



infrastructure de production, qui fonctionne sur une Appliance Oracle. A l'intérieur de nos outillages, nous avons déjà des algorithmes construits par Randstad. Nous travaillons sur le prédictif avec l'aide d'experts, qui nous aident à structurer la démarche mais nous construisons aussi nos propres algorithmes. **Nous utilisons également des algorithmes existants, qui sont déjà établis et validés par la communauté.**

### QUELLE EST LA CONSTITUTION DE L'ÉQUIPE ?

En interne, nous disposons d'une équipe étude, spécialiste des données, d'une équipe IT plutôt spécialiste du développement et nous avons fait émerger un pôle Big Data au sein du pôle Analytique. Nous avons des collaborateurs aux études qui sont plutôt des statisticiens, orientés mathématiques, et des personnes issues de l'IT. **Nous avons fait le choix de ne pas de fusionner les équipes.**

Nous faisons également appel à des profils externes. Quand nous avons besoin, dans notre

roadmap, de monter en charge sur les fonctionnalités à créer, nous nous appuyons notamment sur des ressources de Capgemini.

### PENSEZ-VOUS QUE LE BIG DATA SOIT LA VÉRITABLE SOLUTION À LA PÉNURIE DE RESSOURCES NOTAMMENT DANS LES MÉTIERS TECHNIQUES, TELS QUE LES DATA SCIENTIST ?

**Ce qu'aide à faire le Big Data, c'est de ne pas rester cloisonné dans une recherche standard.** Par exemple, lorsque l'on parle de Big Data ou de Data Scientists, il y a de nombreux métiers recouverts. Il est très difficile de regrouper tous les métiers. La problématique de nos clients est que les demandes sont très formatées. L'un des apports du Big Data, c'est justement d'arriver en questionnement par rapport à ce format, par rapport aux usages, aux recherches, au sourcing et essayer de trouver des solutions éclairées par de l'information et de la donnée, pour arriver à contourner la problématique de la pénurie. Parce que l'on va scénariser la recherche, on va pouvoir prendre en considération une personne un peu éloignée du poste recherché mais que l'on va former, à qui l'on va faire du mentoring. On va également pouvoir envisager de réorganiser l'équipe. Le Big Data dans le secteur de la recherche apporte des réponses à des problématiques que vous pouvez itérer, mais il n'est pas une solution magique.

L'approche par les données aide également à remettre en cause les convictions internes. C'est particulièrement important dans les ressources humaines, car les idées préconçues sont très présentes. Un éclairage par la donnée permet d'arriver à ouvrir un dialogue. **L'approche française des profils est extrêmement standardisée.** Quand vous recherchez quelqu'un, vous recherchez un parcours, une école type. C'est moins le cas dans d'autres pays d'Europe, en Hollande par exemple un niveau d'étude dans n'importe quel domaine démontre une capacité d'ouverture, d'apprentissage, d'investissement et ouvre des portes sur une multitude de postes. **La donnée peut là aussi être intéressante, en rapprochant des savoir-faire ou de savoir-être sont fondamentaux pour un poste.** Cela permet de repenser la manière dont on recrute, dont on fait le sourcing et pense les profils.

### QUELLE ROADMAP ET QUELLES PISTES DE DÉVELOPPEMENT ?

Nous souhaitons introduire du prédictif dans de notre solution. A court terme, nous travaillons sur l'intégration des données de salaire, à l'intérieur de la plateforme pour permettre d'inclure dans les trajectoires métiers les trajectoires de rémunération. Nous voulons enrichir notre plateforme avec de nouveaux jeux de données. La Formation nous intéresse également, car souvent quand on débouche sur une recommandation d'évolution d'un métier A vers un métier B, on constate des écarts de savoir-faire. On pourrait ainsi faire une proposition de type de formation pour pouvoir la compléter. Nous avons créé un comité autour du Big Data, dans lequel nous récupérons des besoins pour créer notre roadmap, dans laquelle les métiers ont une part extrêmement importante.

Enfin en interne, d'autres entités européennes Randstad exprime un intérêt pour les cas d'usage que nous avons développés. Nous élargissons donc la plateforme en Europe. Augmenter le volume de données va nous permettre d'améliorer la pertinence des réponses aux questions et nous donner une vision plus globale à l'échelle des territoires.

### QUE RECOMMANDERIEZ-VOUS AUX PERSONNES QUI SOUHAITENT METTRE EN PLACE UN PROJET BIG DATA ?

Nous sommes partis de cas métiers robustes. **Sans cas métiers robustes, il n'est pas possible de réussir.** La question par laquelle on vient à s'interroger sur l'utilisation et la mise en place d'une solution Big Data doit impérativement être liée à un cas métier, qui apporte de la valeur à l'utilisateur.

Des cas d'attrition de client, d'efficacité commerciale, de prédiction des prochains prospects à approcher par les commerciaux font sens par exemple. Dans certains cas, le Big Data peut ne même pas être utile, il peut être juste suffisant de mettre en place de l'analytique.

**Nous avons commencé par des cas business extrêmement robustes, qui ont donc trouvé du sponsorship fort au sein de l'entreprise.**

Il y a quelques années, les technologies étaient très mouvantes. Aujourd'hui, on dispose de plus de profondeur d'historique, les solutions Big Data sont matures et il est relativement facile de se faire accompagner sur cette problématique.

Sans cas métiers robustes, il n'est pas possible de réussir

# dstad

ORACLE et Capgemini

ANTICIPER

CONNAÎTRE SES CONCURRENTS QUI CHERCHENT LES MÊMES COMPÉTENCES

ÉVOLUER GRÂCE AUX PASSERELLES DE COMPÉTENCES

SE FORMER

ADAPTER LES FORMATIONS

ADAPTER LES POLITIQUES D'EMPLOI

À LA SITUATION

Facilitation visuelle:  
Aline Rollin



## MARK VAN RIJMENAM

Founder

**DATAFLOQ**

*Big Data Strategist and Author of Think Bigger*



# Mixed Data, BDaaS, APIs... Future trends in Big Data

**Mark van Rijmenam** is the founder of Datafloq, the one-stop source for Big Data information and a platform connecting stakeholders within the global Big Data market. He is an entrepreneur, a Big Data strategist and a highly sought-after keynote speaker. He is author of the best-selling book *Think Bigger - Developing a Successful Big Data Strategy for Your Business*, and has been named a global top 10 Big Data influencer. He is currently a PhD candidate studying Big Data and Strategic Innovation at the University of Technology, Sydney.

### *WHY DID YOU CREATE DATAFLOQ?*

The observation I made was that there was little understanding of Big Data and what it can bring to companies, and that lots of organizations were still finding it difficult to get the most out of their Big Data strategies. That was why 4 years ago I decided to set up Datafloq. It was initially called BigData-Startups.com, but we rebranded the firm 2 years ago. The main idea is to connect stakeholders and actors within the Big Data ecosystem, thereby helping them face the global Big Data demand. The objective is to educate the market to drive innovation and economic growth. There are over 150 bloggers creating high-quality content on Big Data trends, privacy, the Internet of Things, technology and offering organizational advice. 5,000 vendors and suppliers are listed and referenced on the website, along with their contact details and locations. The aim is to make them easier to find and to connect with. A recruitment platform has been launched, too. 400 job offers have already been published, and there are many more to come. Last but not least, the platform also advertises key events in the Big Data field. Our aim is to attract all the industry stakeholders and bring them together.

### *WHAT IS THE FUTURE OF DATA SCIENCE AND BIG DATA? SMART DATA, REGARDLESS OF ITS SIZE?*

That is a very broad question. To me, the term 'Big Data' is largely misunderstood. Lots of people think it means they will need a lot of data, resulting in lots of companies asking themselves if Big Data is really for them, as they do not always have large volumes of data.

That's why I like to call it **Mixed Data, which indicates that it is about combining different data sources, regardless of size**. We generate data every day, external and internal data, structured and unstructured data, public data, Twitter data... Combining data sources from a variety of angles is the real challenge and offers the most insights. Companies therefore need to arm themselves with better algorithms to combine this data and derive action from it. Gartner called it the **algorithmic business**. Algorithms are now driving insights from the massive amounts and varying sources of data being generated.

**Mixed Data** is about combining different data sources, regardless of size

### *WHAT ARE THE BOUNDARIES BETWEEN BIG DATA, MACHINE LEARNING AND DEEP LEARNING?*

I think almost all Big Data vendors are using machine learning, and I am seeing more and more algorithms appearing. Companies can even buy algorithms. Algorithmia, for instance, is a marketplace that already exists where you can buy algorithms.



But artificial intelligence is still basic. We saw Google's AI win against Go master Lee Sedol, but in a very specific situation it had been trained for. We still don't have a more generic AI version that could be used in a variety of situations. There is still a long way to go, but progress is being made quite rapidly as we have ever-better computers and processing combined with smarter algorithms.

## EDITOR'S NOTE

### WHAT IS ALGORITHMIA?

This startup founded in 2015 by Diego M. Oppenheimer and Kenny Daniel, gives developers the ability to turn algorithms into scalable web services with a single click. Application developers can integrate the algorithm into their own applications with under 10 lines of code. Algorithmia hosts the web services, makes them discoverable and enables algorithm developers to get paid for usage.

Algorithm developers can host their work on the site and charge a fee per-use to developers who integrate the algorithm into their own work. The platform encourages further additions to its library through a bounty system, which lets users request algorithms that researchers familiar with the field can contribute from their work or develop from scratch for a fee.

More than 800 algorithms are already available on the marketplace, providing the smarts needed to do various tasks in the fields of machine learning, audio and visual processing, and even computer vision.

### ARE WE HEADING TOWARDS AN INCREASING AMOUNT OF AUTOMATION IN DATA SCIENCE AND STATISTICAL MODELLING?

Algorithms will eventually take over many of our jobs - multiple studies indicate that in a few decades' time, **50% of jobs will disappear**. But, of course, humans will always be needed to build the algorithm and the IT infrastructure. Machine learning needs human interaction.

50%  
of **jobs**  
will  
**disappear**

### HOW DO YOU SEE THE FUTURE ROLE OF BIG DATA SCIENTISTS OR CIOS, FOR INSTANCE?

The role of executives will change, because data is becoming more and more important and companies need to become data-driven organizations. **The Chief Data Officer will most definitely have to be present in the boardroom alongside the Chief Security Officer and the Chief Information Officer.**

There is no doubt whatsoever that data governance and data creation must be dealt with in the boardroom, something we can already see happening in the Fortune 500 companies.

What most **companies** are still **missing** is a real **data-driven** company **culture**

### WHAT CONDITIONS STILL NEED TO BE MET FOR A SUCCESSFUL BIG DATA STRATEGY?

I think **what most companies are still missing is a real data-driven company culture**. It is a problem related to people. You can have the latest insights, but without the people to drive it into the business, it is inefficient. Companies need to introduce cultural change management to move to a data-driven and data-centric organization. All employees should have a good understanding of what Big Data is. **I am a big supporter of Big Data and coding in the classroom - teaching future generations about Big Data is absolutely vital.** The subject should already be on the curriculum in primary schools - it is just as important as mathematics or learning another language. Estonia, for instance, has already introduced programs to teach its pupils IT development skills.

**Data governance is another aspect that organizations don't focus on enough.** They should do whatever it takes to make their data secure, because if they fail, they will be hacked and their data will be breached. This can lead to multiple bankruptcies. Four ethical guidelines should be followed:

**TRANSPARENCY:** you have to communicate in a highly transparent fashion;

**SIMPLICITY:** everyone should be able to understand what is being done with their data, both now and in the future;

**PRIVACY:** at all levels - everyone should be building trust through transparency;

**SECURITY:** every organization will be hacked. If you are not, it means you are not that important.



Ethics is truly important; organizations have to treat their data as they would like to be treated themselves. Duckduckgo.com is a search engine that does not store any data about you - it is the opposite of Google. However, organizations need to be aware that they should use data correctly or customers will simply switch to their competitors.

**WHAT WOULD YOU SAY ARE THE BEST EXAMPLES OF SUCCESSFUL BIG DATA STRATEGIES OR THE MOST FORWARD-LOOKING COMPANIES?**

**WalMart is among the best examples.** They were already doing Big Data when most of us were not even considering doing analytics. They collect **40 petabytes of data each day**, combining different Big Data approaches to offer the right customer the right price at the right time and via the right channel.

The health sector is very interesting as well. With electronic health records, England is moving towards harvesting all its medical and health data. There may be privacy issues, but it certainly brings huge opportunities. The Aurora Health Care centre is another successful use case. They have just completed Smart Chart, a \$200 million record system that has accumulated all the **data collected in the past 10 years into a single data warehouse. Data collected from 1.2 million customers, 15 hospitals, 185 clinics, more than 80 community pharmacies, over 30,000 employees including over 6,300 registered nurses, and nearly 1,500 employed physicians.** The not-for-profit Aurora Health Care system has decided to put that wealth of data to good use in order to improve decision-making and make the organization more information-centric. Using electronic and medical data, doctors and DNA data, they have generated a bigger picture of the patient to be able to recommend the right treatment. As a result, admission rates have dropped.

**WHAT ABOUT THE LANDSCAPE OF PROVIDERS?**

Seeing Big Data services offered in the cloud is nothing new. Over the past few years, we have seen many Big Data vendors create Big Data solutions that can be accessed via the web to crunch and analyse your data. More recently, however, we have witnessed the rise of a new type of offering: **Big Data-as-a-Service solutions.** These solutions differ from Software-as-a-Service solutions or Infrastructure-as-a-Service solutions, as they are more or less a combination of the two. This results in a complete package for companies keen to start working with Big Data. Big Data-as-a-Service basically brings together data analytics services for analysing large data sets over the web, while also hosting all that data on scalable cloud hosting services, such as Amazon Web Services. It is therefore a complete Big Data solution, accessible over the web, which doesn't require an in-house solution or a lot of Big Data expertise, thereby enabling small organizations to also benefit from Big Data. We will see more and more BDaaS that enable small companies to plug and play. Bigger companies will require more personalization, but it can help them start proof of concept without needing to invest too much money.

We have witnessed the rise of a new type of offering: Big Data-as-a-Service solutions

**APIs and Applications** are another trend in the Big Data solutions landscape. An **Application Program Interface** is a set of routines, protocols and tools for building software and applications. An API specifies how software components should interact and is used when programming graphical user interface components. A good API makes it easier to develop a program by providing all the building blocks. A programmer then puts the blocks together. APIs will become more and more important, especially for data sets you don't want to own but need to use. The problem is that some major APIs like Twitter are trying to restrict access to their APIs.

Finally, **Data Visualization** will start taking up more space in the landscape, as it allows us to understand what the data is actually telling us. Augmented and Virtual visualization will give the data a whole new meaning. It will immerse us in the data, for instance with a 360° screen to play around with the data.





## OLIVIER EZRATTY

Conseil en Stratégie de l'Innovation



# Les avancées de l'Intelligence Artificielle

**Olivier Ezratty** allie une bonne connaissance des technologies numériques et de leurs applications dans les industries traditionnelles comme les médias, les industries et les services. Il intervient régulièrement dans l'écosystème des startups numériques de par son rôle de président des comités d'agrément de Scientipôle Initiative, ses contributions au sein de Cap Digital et la rédaction du Guide des Startups,



diffusé gratuitement depuis 2005. Veilleur technologique, il s'exprime notamment au travers de son **blog Opinions Libres dédié aux stratégies et politiques de l'innovation**, à l'entrepreneuriat ainsi qu'à une veille technologique multi-facettes. Il rédige notamment un rapport annuel de visite du CES de Las Vegas. En 2015, il a publié **9 chroniques sur les avancées de l'Intelligence Artificielle (IA)** en s'appuyant en grande partie sur une recherche bibliographique extensive, c'est dans ce cadre que nous avons souhaité l'interviewer.

### *POURQUOI VOUS-ÊTES-VOUS INTÉRESSÉ À L'INTELLIGENCE ARTIFICIELLE ?*

Je m'y suis intéressé par curiosité, étant toujours à l'affût de comprendre les technologies émergentes complexes et mal vulgarisées, comme le séquençage du génome et ses applications.

Je rentre dans une démarche pour essayer de comprendre quels en sont les sous-jacentes scientifiques puis les applications marché concrètes. Je souhaite valider si les propos des médias sont factuels. En matière d'Intelligence Artificielle, j'ai même regardé les émissions de Jeopardy que Watson a gagnées pour en comprendre tous les tenants et aboutissants.

### *COMMENT DÉFINIRIEZ-VOUS L'INTELLIGENCE ARTIFICIELLE ?*

A l'origine, il s'agissait de reproduire dans la machine des éléments de raisonnement issus de l'intelligence humaine. Dans la pratique, l'univers de l'IA recouvre plus de choses que l'intelligence humaine. Le traitement des grands volumes de données et l'identification

de tendances dépassent le cerveau humain. L'IA ne reproduit pas pour autant tous les mécanismes de l'intelligence humaine. L'IA a donc déjà dépassé l'intelligence humaine dans certains domaines, et pas dans d'autres. Les deux sont pour l'instant très complémentaires.

### *EN QUOI L'INTELLIGENCE ARTIFICIELLE SE DIFFÉRENCIE-T-ELLE DU BIG DATA ?*

Dans le principe, le Big Data relève de projets d'exploitation de grands volumes de données.





Les techniques utilisées, de type Machine Learning, s'apparentent au fonctionnement du cerveau pour identifier des patterns, des tendances ou des signaux faibles. La segmentation, l'étude des données comportementales et socio-économiques sont difficiles à faire avec des techniques statistiques traditionnelles. Les statistiques, basées sur les modèles bayésiens, consistent en l'interrogation de bases de données relativement classiques.

En revanche les techniques de traitement du langage, Natural Language Processing (NLP), utilisées pour exploiter les données, sont déjà à la frontière de l'Intelligence Artificielle.

Les "victoires" de l'IA, Watson dans Jeopardy en 2011 et AlphaGo en 2016, s'appuient sur l'exploitation de très gros volumes de données. Même si ces progrès peuvent paraître impressionnants, il s'agit d'évolutions modestes car elles sont réalisées dans des domaines très spécialisés. Watson, s'appuie également sur de larges volumes de données pour aider les cancérologues à adapter les traitements des patients. **L'Intelligence Artificielle manifeste sa puissance quand elle a accès à beaucoup de données.**

**L'Intelligence Artificielle**  
manifeste sa **puissance**  
quand elle a accès  
à **beaucoup de données**

Dans la reconnaissance de la parole, les techniques nécessitent au départ un long apprentissage. Il fallait répéter des mots pour permettre l'apprentissage du système. Ces techniques évoluent pour qu'il n'y ait pas besoin d'apprentissage spécifique à chaque utilisateur.

Dans la plupart des cas, les avancées dans le domaine de la reconnaissance de la parole sont issues de la recherche publique, de recherche et développement au sein de grandes entreprises comme Amazon, Google ou encore Apple, mais elles proviennent également d'acquisition de startups.

Apple utilise notamment les technologies de reconnaissance de la parole de Nuance dans son logiciel Siri. Le marché de la reconnaissance vocale s'est consolidé, l'américain Nuance ayant notamment racheté les droits de la société belge Lernout & Hauspie ainsi que Dragon Voice recognition, qui était distribué par IBM.

Les recherches portent désormais plus sur les agents conversationnels type chatbot, qui peuvent s'interfacer avec la reconnaissance de la parole. C'est ce que fait actuellement Google Home.

**DANS L'ENSEMBLE DES CLASSIFICATIONS RECENSÉES, LAQUELLE VOUS PARAÎT ÊTRE LA PLUS PERTINENTE ?**

Les classifications à très haut niveau de l'Intelligence Artificielle relèvent de la science-fiction.

**L'ARTIFICIAL NARROW INTELLIGENCE (ANI)** correspond à l'état de l'art actuel dans le domaine : des solutions d'IA très spécialisées. Cela a commencé avec les systèmes jouant et gagnant aux échecs comme Deep Blue d'IBM en 1997, puis avec des systèmes experts pointus comme dans certains secteurs de la santé. On peut y classer les moteurs de recherche courants, la détection de fraudes bancaires, le credit rating de particuliers, la conduite automatique ou assistée, Apple SIRI, Microsoft Cortana et Google Translate.

**L'ARTIFICIAL GENERAL INTELLIGENCE (AGI)**, est celle qui se rapproche le plus de l'intelligence humaine avec un côté polyvalent et notamment la capacité à raisonner, analyser des données, résoudre des problèmes variés et notamment raisonner par analogies. On peut intégrer dans ce niveau un grand nombre des capacités humaines : l'usage du langage, de la vue et des autres sens, la mémoire, la pensée, le jugement et la prise de

décisions, la résolution de problèmes multi-facettes, l'apprentissage par la lecture ou l'expérience, la perception du monde et de soi-même, l'invention et la créativité, la capacité à réagir à l'imprévu dans un environnement complexe ou encore la capacité d'anticipation, cette liste est très longue ! Pour l'instant, on en est encore loin, même si certaines de ces capacités notamment linguistiques et de raisonnement général sont en train de voir le jour dans l'IA.



**L'ARTIFICIAL SUPERIOR INTELLIGENCE (ASI)** est quant à elle caractérisée par la puissance des machines qui se démultiplie et se distribue plus facilement que celle d'un cerveau humain. A ce niveau, l'intelligence de la machine dépasse mécaniquement celle de l'homme dans tous les domaines y compris dans la créativité et même dans l'agilité sociale.

Dans certains domaines où la puissance brute compte, nous nous trouvons déjà dans l'AGI voir l'ASI. Dans ceux de la reconnaissance de l'environnement, de la reconnaissance visuelle et de l'intelligence émotionnelle, l'Intelligence Artificielle n'égale pas encore celle de l'homme. **On continue de découvrir pas à pas la manière dont le cerveau fonctionne en termes de capacité brute, la manière dont les neurones enregistrent l'information et s'interconnectent. A chaque découverte, cela retarde l'échéance théorique d'une machine qui aurait la sophistication d'un cerveau humain.** Il n'y aura probablement jamais d'Intelligence Artificielle équivalente à l'humain avec la même expérience à l'âge adulte, sa mortalité, son fonctionnement hormonal, ses sens. En revanche les machines nous dépassent déjà dans la capacité d'analyse de très grands volumes de données ne nécessitant pas d'intelligence sensorielle ou émotionnelle. L'Intelligence Artificielle sera parfaite mais sous une forme différente de l'homme. Ce qui obscurcit actuellement notre compréhension du sujet, c'est la définition anthropomorphe qu'on lui donne. Mais nos créations ne sont pas à notre image, n'ont pas nos sens par exemple.

L'histoire des technologies a toujours montré que les nouveautés technologiques généraient de nouvelles peurs comme l'invention du nucléaire, qui engendre pourtant moins de morts que le charbon. Ces peurs sont justifiées en partie mais ne doivent pas limiter nos créations.

Ce qui **obscurcit** actuellement notre **compréhension** de l'IA, c'est la définition **anthropomorphique** qu'on lui donne

### COMMENT ÉVOLUE L'ÉCOSYSTÈME EN TERMES D'ACTEURS ?

L'écosystème de l'Intelligence Artificielle est plutôt simple à comprendre. Il s'agit d'un ensemble de nouvelles techniques avec des usages variés dans de nombreux domaines au même type que le cloud par exemple. L'Intelligence Artificielle est intégrée dans un grand nombre de startups technologiques. On voit naître des Pure Players avec de nouveaux usages basés uniquement sur des techniques d'Intelligence Artificielle, des briques open source qui utilisent des sources de données variées, qui sont assemblées à la main. Dans la plupart des cas, nous sommes sur des démarches empiriques et expérimentales.

**On exagère parfois l'avance des grands acteurs de l'Intelligence Artificielle.** Google dénombre beaucoup de spécialistes de l'Intelligence Artificielle, mais la masse critique des chercheurs est dans les laboratoires publics. Ils sont d'ailleurs souvent amenés à fonder leurs propres startups. DeepMind, créé en 2010 et acquis par Google en 2014, était une entreprise de 50 personnes, créé par des chercheurs d'Oxford et de Cambridge, qui ont appliqué les résultats de leurs travaux de recherche.

Aux États Unis, les produits des travaux des chercheurs de Stanford, du MIT sont majoritairement rachetés par Google. Il s'agit d'un phénomène classique qui régit le fonctionnement de l'ensemble de l'écosystème numérique.

Il y a cependant encore beaucoup de chercheurs dans les universités, les laboratoires de recherche publics financés par des deniers publics. **Aux États-Unis, 90% de la recherche vient des aides fédérales et la moitié est financée par le domaine militaire et les services de renseignements.** L'«Agence pour les projets de recherche avancée de défense» (DARPA) finance la plupart des projets de recherche en robotique. Les crédits DARPA ont par exemple financé Boston Dynamics, que Google est en train d'essayer de revendre.

### SI VOUS DEVIEZ NOUS CITER QUELQUES STARTUPS RECENSÉES DANS VOS CHRONIQUES, QUELLES SERAIENT-ELLES ?

La plus grande moitié des startups en Intelligence Artificielle s'intéresse au secteur du commerce en ligne, au Data Marketing pour prédire les comportements des clients mais les startups s'illustrant dans le domaine de la santé sont, selon moi, les plus prometteuses et positives dans la mesure où elles touchent un sujet sensible et noble.

**CARDIOLOGS** a développé un premier service d'analyse des électrocardiogrammes en ligne, basé sur des algorithmes de Machine Learning, qui fournit aux praticiens des informations pour les aider dans l'interprétation des ECG.

**DEEP GENOMICS** a créé le DG Engine qui analyse les variations du génome, les mutations de l'ADN, et la manière dont elles affectent le fonctionnement des cellules et génèrent des pathologies.



**ENLITIC** propose de l'aide au diagnostic en s'appuyant principalement sur les résultats de systèmes d'imagerie médicale (IRM, scanner, radios) et sur du deep learning. C'est une sorte d'équivalent apparemment généraliste d'IBM Watson qui se positionne plutôt dans la prévention, détectant des pathologies émergentes le plus tôt possible, notamment les cancers du poumon.

**GINGER.IO** a conçu un outil de diagnostic et de prescription de traitement pour diverses pathologies neuropsychologiques. Il exploite des applications mobiles pour le diagnostic et du Machine Learning. La solution permet un auto-traitement de certaines pathologies par les patients.

**LUMIATA** est dans la même lignée, un système d'analyse de situation de patient permettant d'accélérer les diagnostics, notamment en milieu hospitalier.

**BEHOLD.AI** a développé une solution d'analyse d'imagerie médicale pour aider les radiologues à faire leur diagnostic qui s'appuie sur du machine learning. Le système compare les images de radiologie avec et sans pathologies pour détecter les zones à problèmes, comme les nodules et autres formes de lésions.



**COGNITIVE SCALE** a créé la solution Cognitive Clouds. Elle est notamment proposée aux adolescents atteints de diabète type 1 pour les aider à se réguler, en intégrant les aspects médicaux (prise d'insuline, suivi de glycémie), d'activité physique et d'alimentation.

**ATOMWISE** utilise le Machine Learning pour découvrir de nouveaux médicaments et vérifier leur non toxicité. Le principe consiste à simuler l'interaction entre des milliers de médicaments connus et une pathologie telle qu'un virus, et d'identifier celles qui pourraient avoir un effet par simulation des interactions moléculaires. Un premier résultat aurait été obtenu en 2015 sur un virus d'Ebola.

**MEDAWARE** fournit une solution qui permet d'éviter les erreurs de prescription médicamenteuse en temps réel pour les médecins en incluant des briques de Big Data et de Machine Learning qui exploitent notamment des bases de données médicales d'historiques de patients.

### QUELLES SONT LES AUTRES APPLICATIONS DE L'INTELLIGENCE ARTIFICIELLE ?

On trouve de l'Intelligence Artificielle appliquée aux moteurs d'analyses prédictives, à la recherche visuelle avec des applications sectorielles diverses. Les applications de l'Intelligence Artificielle dans les services financiers sont nombreuses avec de l'optimisation de taux d'intérêts de prêts, de la détection de fraude, du credit rating d'emprunteurs basé sur les réseaux sociaux, de l'optimisation de planification financière et d'investissements. Elles sont également nombreuses dans le secteur du commerce, du marketing, des ressources humaines et des services juridiques. L'Intelligence Artificielle se voit également appliquée dans la sécurité informatique, l'agriculture et bien sûr la recherche scientifique. **Kaggle est par exemple une communauté mondiale de Data Scientists travaillant sur des défis d'Intelligence Artificielle.** L'Intelligence Artificielle peut permettre d'augmenter la productivité dans tous les métiers qui manipulent un corpus de données très dense, que le cerveau humain n'est pas en mesure de mémoriser en entier.

### QUELLE EST LA MATURITÉ DES ENTREPRISES FRANÇAISES EN TERMES D'IA ?

Je n'observe pas d'avance ou de retard particulier en France. Les recherches sont très actives au sein de l'INRIA et du CNRS desquels émergent de nombreuses startups. L'Europe dispose d'une tradition scientifique et mathématique particulièrement intéressante pour les recherches en Intelligence Artificielle. Beaucoup d'entreprises utilisent l'Intelligence Artificielle pour perfectionner leurs offres ainsi que comme outil de communication, une formule pour surfer sur la vague actuelle. **Nous accusons cependant toujours un certain retard dans l'internationalisation de nos startups, dans l'IA comme dans tous les autres secteurs.**



## FRANCISCO DE SOUSA WEBBER

Co-Founder and Inventor

**CORTICAL.IO**



# Computable Language & Natural Language Processing

**Francisco De Sousa Webber** first took an interest in Information Technology as a medical student specializing in genetics and serology at the University of Vienna. He participated in various research projects at the Vienna Serological Institute and was heavily involved in medical data processing. He was also involved in numerous projects, including establishing and organizing Austria's dialysis register database and creating a patient documentation system for the university clinic.

In the mid-1990s, he worked alongside Konrad Becker to found Vienna's Institute for New Culture Technologies and Public Netbase - Austria's only free public-access Internet server at the time - thus establishing an international competency platform for the critical use of information and communication technologies.

In 2005, Francisco founded Matrixware Information Services, a company that developed the first standardized database of patents under the name of Alexandria, where he acted as a CEO. He also initiated the foundation of the Information Retrieval Facility, a not-for-profit research institute, with the goal of reducing the gap between science and industry.

He currently heads up Cortical.io, a start-up he co-founded in 2011 that develops and commercializes Natural Language Processing (NLP) solutions based on Semantic Folding, a theory that offers a fundamentally new approach to handling Big Text data.



## WHY DID YOU CREATE CORTICAL.IO, WHAT WERE YOUR MOTIVATIONS?

As part of my research work, I was handling data and statistics. Things changed over time and I found I was struggling to understand what the doctors were saying about patients. I wanted to understand how we could go about extracting the exact meaning from the texts they wrote. Everything is clear with numbers, but far more unclear when it comes to texts. I therefore started to specialize in text analysis. When I was developing search engine technologies, I worked with numerous large companies to create applications, notably in the field of patents. **I soon realized that this sector had very little access to high-performance technologies and that there was very little progress in developing technologies to treat complex information such as technological and legal languages.**

In 2005, I therefore decided to start a new company specifically in this field. We had been working with research corporations and universities to study natural language processing in great detail, but the approach they had was too far removed from what people were actually doing. We consequently decided to sell Matrixware Information Services to a US partner.

I began looking for an alternative and statistics-free approach to Natural Language Processing (NLP). My conclusions were that the only system that has proper Natural Language Processing capabilities is the human brain. Our approach therefore needed to be as similar as possible to the human brain. That was how Cortical.io came about.



## WHERE DID YOUR INSPIRATION COME FROM? DID ANY WORK IN PARTICULAR INSPIRE YOU?

**My main inspiration is, of course, how the brain processes information.** I found key inspiration in this field in 2005 with **Jeff Hawkins' theory about how the neocortex works**, and began my own experimental work in 2010 to see whether his theoretical framework could be applied to language. Luckily, we were able to get a research grant from the Austrian Research Promotion Agency - FFG Science in Austria. This grant enabled us to hire a team of computer scientists in order to develop a prototype. Half way through the experimentation phase, the results were so astonishing that we had to accelerate the process and expose our findings to the market to test it. We found an angel investor in 2012 and started out on the journey of being a start-up in Natural Language Processing.

## HOW DOES THE RETINA WORK? WHAT IS A FINGERPRINT, EXACTLY?

We have developed a new machine learning approach inspired by the latest findings on the way the brain processes information. Our approach uses similarities as a foundation for intelligence. **By mimicking the understanding process of the brain, we benefit from millions of years of evolutionary engineering to help us solve today's hottest NLP challenges.**

Our system learns in a similar way to the way the brain works when we perceive words. **When a child learns a new word, he stores every possible utterance on a mental map where similar meanings are organized in close proximity.** For example, the different meanings associated with the word 'organ' (music, church, liver...) are each stored in specific places in this mental map. Depending on the context, our brain automatically associates the word 'organ' to the correct semantic cluster – this is what we try to reproduce with our Retina Engine.

Cortical.io's Retina Engine learns about a specific language by processing relevant text content via unsupervised learning. **It converts words into semantic fingerprints, a data format similar to the one used by the brain that captures the meaning behind natural language.** The Retina can generate semantic fingerprints for different language elements such as words, sentences and entire documents.

A semantic fingerprint is structured like a map where you can visualize the overlap between words or sentences. This type of data representation makes it very easy to compare any two

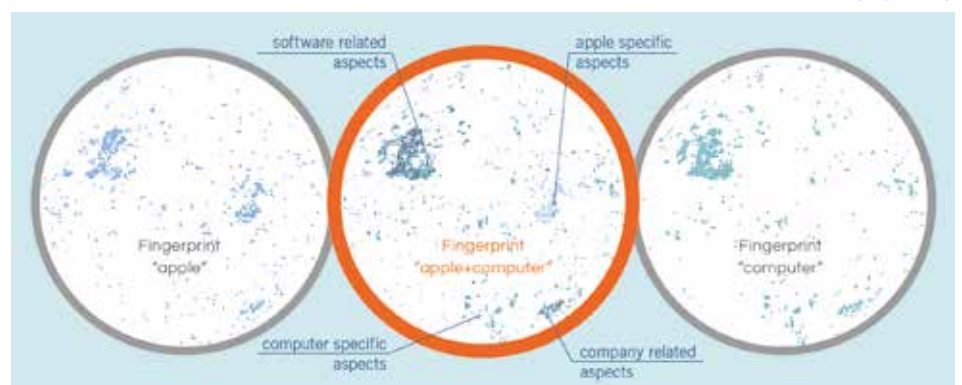
words, sentences or even whole documents, because you can calculate their similarity and measure their semantic overlap.

By mimicking the understanding process of the brain, we benefit from millions of years of evolutionary engineering to help us solve today's hottest NLP challenges.

A concrete example would be a wildlife journalist who is looking for information on the Web about jaguars. No matter how many newsfeeds and keywords he has selected, he will always receive tons of non-relevant information about Jaguar, the car, rather than solely information about the animal. This is because computers only look for keywords in an article, without understanding the meaning behind them. In this example, Cortical.io's Retina engine enables an intelligent news filter to be created based on semantic fingerprints. First of all, the journalist's interests are captured in a filter fingerprint. Then, every article is converted into a semantic fingerprint. Finally, the Retina compares each article's fingerprint with the filter fingerprint, and the system forwards only the articles that are extremely similar to the journalist's filter fingerprint.

**What is very interesting is that, with our approach, semantic spaces are stable across languages:** the semantic fingerprints of the word 'philosophy' in Chinese, French, English or German (for example) all look very similar. This means that the Retina Engine makes it possible to directly compare documents that have been written in different languages, or, for example, to search a database of Chinese documents using English words.

Semantic Fingerprinting



## WHAT ARE THE DIFFERENT POSSIBLE USES OF SEMANTIC FOLDING?

Semantic Folding can have various applications and be used in different fields, including:

**SOCIAL MEDIA:** the semantic fingerprint of the description of a person can be compared with another profile, for instance. Semantic Folding can also be used for social content streaming or to detect abnormal behaviour in social media.

**BANKING AND COMPLIANCE MONITORING:** Banks have to monitor their communications for things such as inside trading. Previously, no satisfactory email checking solutions existed, because metaphors could be used in emails to cover potential fraud. Converting the messages into fingerprints has allowed banks to detect more frauds and to reduce false positives.

**CONTENT MANAGEMENT:** when creating a website, its Google ranking will depend on the quality of the information. Your content has to be interesting to be ranked among the top websites. With the Retina Engine, you can create the fingerprints of the 20 top Google-ranked pages in your particular category. By overlapping those fingerprints with the fingerprint of your own content, you can see which parts of your text you should adjust to improve your website's visibility.

**AUTHOR DETECTION:** say you want to automatically associate new publications with the corresponding author. With the Retina Engine, you simply need to create a fingerprint of one of the author's most typical articles and then compare the overlap of any new publication with that reference fingerprint to associate it with the corresponding author. This is particularly useful for digital publishers.

There are many examples of other uses, too, ranging from the analysis of TV captions to determine the best topics for shows to terrorism prevention and reputation management...

## HOW DO THINGS CURRENTLY STAND AS REGARDS ARTIFICIAL INTELLIGENCE? DO YOU THINK BUSINESSES ARE READY FOR IT?

Austria is definitely not ready for this kind of technology, but **Silicon Valley and New York have already thoroughly embraced it.** Generally speaking, lots of customers say they have been trying Machine Learning solutions but that nothing suits their business activity and that it isn't really working. They say that big software providers could not handle their requirements. As a matter of fact, most businesses are not aware of the solutions that are out there; everybody is talking about Machine Learning, but not about Natural Language Processing. **There is a big gap between academic research and what is actually happening in companies, even for big players. Companies are still gathering Tera octets of data without having any idea of what to do with it.**

Our advantage is that Cortical.io is not a black box, and people understand our approach quite easily.



## WHAT DO YOU THINK THE FUTURE HOLDS FOR ARTIFICIAL INTELLIGENCE?

In my opinion, AI is not going to take over the world. Theoretically this could happen, but considering the path of our history, I think it is highly unlikely.

The brain is a specific size and the size of the near cortex is limited, so **I believe it will all be about adding intelligence to it. I believe we will continuously create larger patches and extend our individual cortex.** We will be able to add a cortex if we want to detect messages in Japanese, for instance. People will extend their personality with a kind of added exoskeleton.

I don't feel threatened by autonomous Artificial Intelligence, but more by people who could extend their capabilities and the ethical authority of the person wearing the extension.

**However, high-level autonomous AIs are not the first things we are going to see, we will begin by empowering individuals.**

There is a big gap between **academic research** and what is **actually happening in companies.** Companies are still gathering **Tera octets of data** without having any idea of **what to do with it**





## CAROLINE CHOPINAUD

CBDO

craft ai

## CLODÉRIC MARS

CTO

craft ai

# craft ai: L'Intelligence Artificielle as-a-service

**Caroline Chopinaud** a fait des études d'informatique et a un doctorat en Intelligence Artificielle de Paris 6. Elle s'est spécialisée sur les systèmes multi-agents et agents intelligents etc... Elle a été enseignante à l'Université pendant deux ans avant de débiter chez MASA, la maison mère de craft ai. Elle a débuté en tant que Développeuse sur leur solution d'Intelligence Artificielle appliquée à la modélisation de comportements dans le domaine de la simulation d'entraînement pour la défense et la sécurité civile. Elle a ensuite pris le poste de Directrice R&D de la société pendant 3 ans, avant de devenir Responsable du développement commercial et des partenariats de craft ai.

**Clodéric Mars** est Ingénieur de l'INSA de Rennes. Il a travaillé au sein d'un laboratoire de l'INRIA sur des projets d'Intelligence Artificielle appliquée aux personnages de jeux vidéo et de simulations, réunissant des aspects 3D et IA. Il a travaillé pendant trois ans dans une société de services sur la création de villes en 3D et sur toute la chaîne d'outils de création. Il a ensuite rejoint Golaem, spin-off de l'INRIA. Il a travaillé sur des problématiques de simulations de foule depuis l'animation jusqu'à l'Intelligence Artificielle des personnages, mises en œuvre pour des effets spéciaux de films d'animation. Il a rejoint MASA en tant que responsable technique du produit MASA LIFE, middleware d'IA pour les jeux vidéo et la simulation, qui a servi de fondation à la création de craft ai. Après 3 ans à travailler sur ce produit, il a pris le rôle de CTO au sein de craft ai.

## QU'EST-CE QUI VOUS A POUSSÉ À CRÉER craft ai ?

La naissance de craft ai est partie d'une volonté de MASA Group de diversifier ses activités. Cette diversification avait débuté via le middleware d'IA MASA Life, appliqué aux jeux vidéo, et en R&D où des projets de jeux et de robotique avaient vu le jour. L'objectif était de prototyper et d'expérimenter nos solutions d'IA à d'autres secteurs. Ainsi, fin 2014, nous avons développé un premier prototype promoteur d'assistant personnel fondé sur notre technologie MASA Life, sur lequel nous avons expérimenté des approches d'automatisation de gestion de rendez-vous en fonction des préférences utilisateurs, d'assistant CRM et de dialogue d'interaction simple. Nous nous sommes rapidement rendu compte qu'il y avait de nombreuses applications possibles de notre technologie et notre savoir-faire en IA, qui allaient bien au-delà des jeux et de la simulation. C'est à partir de ces premières études et expérimentations que nous avons commencé à officiellement travailler sur le projet craft ai début 2015 et exploré différents marchés.

Cette prospection nous a permis de mettre en évidence le fait que nos technologies répondaient à un besoin bien précis pour lequel peu de solutions existent actuellement sur le marché. Sur l'année 2015, nous sommes essentiellement allés sonder le marché, et notre approche faisait écho dans le monde des assistants personnels, des bots, de la robotique mais également des applications mobiles et surtout dans l'IoT. En effet, pour tous ces domaines, **la prise en compte de l'utilisateur final en tant qu'individu unique est particulièrement importante**. Les services proposés doivent être capable de s'adapter en continu à chaque utilisateur ! C'est ce que propose craft ai. **Une solution horizontale à destination des développeurs pour faciliter la création de services personnalisés.**

## POUVEZ-VOUS NOUS PARLER DES ÉQUIPES craft ai ?

Nous sommes une équipe de 12 personnes avec une majorité de profils ingénieurs en informatique et docteurs en Intelligence Artificielle. Initialement nous sommes pour la plupart des développeurs et des chercheurs, issus de MASA. Nous avons une expérience de plusieurs années en développement de solutions d'IA et d'outils de développement simples d'utilisation, appliqués à des domaines complexes. Nous préférons les profils alliant théorie et pratique. **Nous n'avons ainsi pas, à ce jour, de profils types « Data Scientists », juste de très bons développeurs qui maîtrisent les aspects théoriques de l'IA. Ce sont des profils très difficiles à recruter !** Nous sommes toujours à la recherche de jeunes docteurs en Intelligence Artificielle mais il y a beaucoup de concurrence en termes de recrutement sur Paris. Il est compliqué pour une entreprise en développement de concurrencer les laboratoires de grandes entreprises comme FAIR, ou encore SNIPS. C'est une difficulté globale dans la Tech, les bons profils sont rares et demandés, mais la manière dont nous sommes organisés implique que tout le monde ait un impact très fort sur notre produit. **Les personnes qui nous rejoignent ont une sensibilité tournée vers les utilisateurs et souhaitent voir des projets émerger des outils qu'elles créent.**

## QUELS SONT LES SERVICES D'INTELLIGENCE ARTIFICIELLE QUE VOUS PROPOSEZ ET QUELS DOMAINES SONT TOUCHÉS ?

Nous développons des solutions d'IA à destination des développeurs, pour leur permettre de concevoir des applications et services qui vont s'adapter à chacun de leurs utilisateurs. Nous mettons à leur disposition une API, avec un algorithme de Machine Learning, qui est axée sur l'apprentissage des habitudes d'un individu en vue d'automatiser, de suggérer, de lever des alertes ou de créer tous types de bots personnalisés. Jusqu'à présent, la plupart des approches de traitement des données utilisateurs fonctionnaient sur du traitement Big Data, qui permet d'analyser et de prendre une décision à l'échelle d'une population. La personnalisation des services fondée sur des approches telles que la recommandation a son intérêt mais ne remplit pas nécessairement toutes les attentes des utilisateurs finaux et des développeurs, même de services, qui souhaitent aller plus loin dans l'adaptation à l'utilisateur. C'est pour cela que **nous appliquons notre approche de Machine Learning à l'échelle de chaque utilisateur. Les vrais challenges dans ce contexte sont sur le peu de données accessibles et le besoin de convergence rapide de l'apprentissage**. Quand nous échangeons avec des banques, des assurances, des sociétés de services qui font déjà du Big Data, de l'analyse statistique et de la segmentation, nous comprenons que la valeur actuelle recherchée réside véritablement dans ce qui est (hyper)-personnalisé et très proche de l'utilisateur.

## EST-CE QUE VOUS POUVEZ NOUS DONNER DES EXEMPLES D'APPLICATIONS MÉTIERS DE craft ai ?

Les secteurs d'application sont nombreux. Dans le domaine de l'IoT et des applications Smart Home, nous intervenons sur **l'automatisation du comportement des objets de la maison**, comme des thermostats par exemple. Nous sommes régulièrement sollicités sur l'apprentissage et l'automatisation de planning et de consignes de chauffe afin d'y intégrer les habitudes des individus.



**Nous menons également des projets liés aux applications mobiles et à l'expérience utilisateurs.** Nous avons pour exemple mené un projet avec le LCL, dont le but était de permettre **l'activation de fonctionnalités qui ne sont pas ou peu utilisées par les clients de leur application mobile**, par difficulté de prise en main ou par peur, telle que l'option d'affichage du solde en « un clic ». Grâce à craft ai, nous avons pu apprendre les habitudes d'un utilisateur de l'application, en particulier les contextes dans lesquels cet utilisateur regarde son solde en fonction d'informations telles que la localisation, l'heure de la journée, le jour de la semaine... Cet apprentissage permet de suggérer, au bout de quelques jours, d'activer la

fonctionnalité d'affichage de solde automatique uniquement dans des contextes bien précis. C'est une forme de « push personnalisé » qui est envoyé au bon moment et pour une bonne raison.

Nous travaillons également sur du **traitement de flux d'information**. A partir de flux RSS d'informations, il s'agit d'apprendre à quel moment une personne regarde telle ou telle catégorie d'information, pour être capable de lui « pusher » de manière proactive l'information qui va certainement l'intéresser et ce au moment le plus pertinent.

Nous mettons aussi notre modèle au service de la **détection d'anomalie dans les comportements** notamment sur le sujet du maintien des personnes âgées à domicile. L'idée ici est d'apprendre les comportements de la personne, grâce aux objets connectés et capteurs présents dans son environnement, pour permettre d'analyser des dérives sur son comportement.



Enfin, nous travaillons sur **le sujet des bots conversationnels**, pas dans la partie langage mais dans la partie adaptation de la manière de répondre à l'utilisateur : le moment, le contenu. Nous n'avons pas été spécialement proactifs sur ces sujets, mais sommes plutôt directement sollicités pour ajouter cette dose de personnalisation, d'interaction, cette couche supplémentaire clef dans ce genre de produit conversationnel. **Cette personnalisation de l'interaction a des applications qui vont au-delà des simples chatbots, vers les assistants personnels et la robotique domestique.**

L'API sera en accès ouvert et complet sous peu. **Notre objectif est d'avoir un modèle d'API libre-service de masse comme STRIPE, TWILIO ou encore ALGOLIA.** Notre modèle n'est pas celui d'une plateforme, comme IBM BlueMix par exemple, mais d'une API qui peut être utilisée dans tous types d'environnements ou plateformes et en combinaison avec d'autres API. En mettant nos technologies dans les mains des développeurs, nous découvrons de nouvelles applications, c'est pourquoi **nous nous efforçons d'aller au-devant des communautés de développeurs au travers de Hackatons**, par exemple.

## EN QUOI VOTRE MANIÈRE DE TRAITER LES DONNÉES SE DIFFÉRENCIE-T-ELLE DU BIG DATA ?

Nous avons développé une approche qui vient de notre passé de jeux vidéo et de simulation qui est fondée sur une approche « agent » qui pour simplifier représente un utilisateur, et qui va vivre avec l'utilisateur et s'adapter à lui en fonction des données qui lui seront fournies. Nous faisons tourner des algorithmes de classification sur ces données en continu pour générer un modèle de décision sous la forme d'un arbre de décision qui est interprété par l'agent pour obtenir un comportement perti-

nent par rapport au contexte et à l'utilisateur. Concrètement si l'on branche un thermostat qui utilise craft ai, nous récoltons assez de données en une journée pour permettre d'avoir un planning dès le lendemain, qui fonctionne avec un bon degré de certitudes et qui va s'affiner au fur et à mesure.

Notre **objectif** est  
d'avoir **un modèle**  
**d'API libre-service**  
de masse comme  
**STRIPE, TWILIO** ou  
encore **ALGOLIA**

**L'arbre de décision est un modèle de décision qui, en plus d'être interprétable par la machine, peut être lu et debuggé par un être humain.** Il peut l'interpréter, le modifier et raisonner sur cet arbre de décision. Nous avons cherché à développer ce modèle pour se différencier des approches plus boîtes noires, comme les réseaux de neurones, car notre objectif est de faire en sorte que notre IA soit accessible à tous les développeurs, et qu'ils puissent comprendre pourquoi la décision a été prise et la contredire éventuellement.

**Nous ne faisons donc pas de Big Data mais nous pouvons travailler avec des entreprises, qui elles amassent un volume important de données.** En effet, si une entreprise a 100 000 clients, cela correspond à **peu de données par client, que nous traitons individuellement et en quasi temps réel.** Le modèle de décision va pouvoir être mis à jour tous les jours et va évoluer automatiquement au fur et à mesure de la vie de l'utilisateur. **C'est le processus inverse du Big Data qui part de toutes les données de la population pour créer un modèle global et potentiellement le descendre sur des segments ou des sous-populations.**

Nous avons des clients qui nous demandent s'il est possible d'utiliser l'API dans un contexte plus large. Par exemple, si un agent ne sait pas prendre une décision dans un contexte donné, mais qu'un autre agent sait prendre cette décision, on se pose la question de savoir comment déterminer et réutiliser la connaissance de cet agent dans l'arbre de décision de l'autre agent. Il est très important de pouvoir apporter une solution à ce type de problème, et il s'agit d'une de nos problématiques de R&D actuelles.

### **SELON VOUS, OÙ EN SOMMES-NOUS DE L'INTELLIGENCE ARTIFICIELLE ?**

**Tant que cela s'appelle de l'Intelligence Artificielle, c'est toujours de la R&D.** En général, on dit en IA que ce qui n'est plus du secteur de la recherche change d'appellation. C'est ce que l'on appelle le « AI effect », une fois que l'IA a résolu un problème, ce n'est plus de l'IA. Nombreux sont les domaines qui font partie

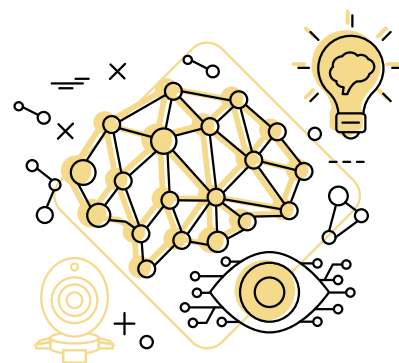
de notre quotidien et ne sont plus considérés comme de l'IA, mais qui il y a 20 ans animaient les laboratoires : recherche de texte, reconnaissances de visages, Knowledge Graph... **Le traitement automatique du langage naturel est également en phase de quitter la sphère de la recherche en IA.**

### **QUID DE LA MATURITÉ DES ENTREPRISES ?**

Les entreprises sont plus matures qu'avant. Il y a encore quelques années de ça, les entreprises avaient peur du terme IA. Aujourd'hui, nous faisons des RDV, durant lesquels les clients nous demandent ce qu'ils pourraient faire avec de l'IA sur la base de notre modèle. C'est plus complexe, car ils n'ont pas de problématique précise à résoudre, mais pensent que l'IA est une réelle plus-value sans savoir comment l'appliquer et pourquoi. **Les entreprises sont revenues de la mode du Big Data, elles mettent un véritable mur entre le Big Data et l'IA alors qu'il y a globalement une continuité entre les deux.**

Enfin, il y a actuellement beaucoup de promesses faites, notamment concernant les assistants dits virtuels, les voitures autonomes, mais cela ne va pas arriver aussi vite qu'on peut le penser, pour des questions législatives notamment.

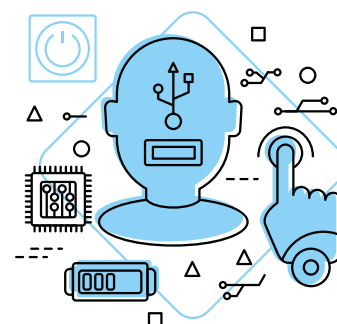
Les entreprises mettent  
**un véritable mur** entre le  
**Big Data et l'I.A.** alors  
qu'il y a globalement une  
**continuité** entre les deux



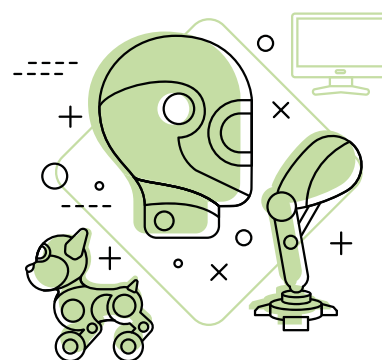
## **Intelligence Artificielle**



## **APIs**



## **Bots Conversationnels**



## **Robotique Domestique**



BÉATRICE  
**TOURVIEILLE**

Directrice Marketing

**VOYAGES-SNCF.COM**



Après plusieurs années d'expérience dans le conseil, **Béatrice Tourvieille** a évolué dans le domaine digital, au sein de Meetic, Deezer et Canal+, toujours dans des fonctions marketing avec une orientation data et connaissance client. Elle a rejoint Voyages-sncf.com au poste de Directrice Marketing, qui regroupe l'acquisition, les achats médias online, le marketing relationnel, la data et la connaissance client ainsi que la partie éditoriale et production de contenus autour des destinations.

**Voyages-sncf.com** :  
Digital et Big Data  
au **service** de  
la **personnalisation**

### QUELLES MOTIVATIONS POUR REJOINDRE LA DIRECTION MARKETING DE VOYAGES-SNCF.COM ?

Ce qui m'a intéressée avant tout c'est l'importance des enjeux et la variété des problématiques proposées dans ce poste. Le Marketing Groupe intervient sur plusieurs marchés, à la fois sur le marché français, sur lequel nous sommes leaders avec de gros enjeux de personnalisation et de fidélisation, mais également sur des marchés internationaux, avec de forts enjeux de conquête. Mes motivations pour rejoindre Voyages-sncf.com résidaient dans l'importance portée à la connaissance client et à la data, et dans la volonté du groupe d'insuffler cette culture sur les canaux marketing et plus globalement dans l'organisation.

Le marché évolue rapidement et se complexifie, avec une offre de modes de transport et de distribution de plus en plus large. Il s'agit d'un enjeu fort pour le développement de notre offre, avec par exemple un nouveau positionnement sur le bus, ou encore la mise en avant de nouvelles offres de transports ferroviaires avec des services permettant de voyager au meilleur prix.

Nous collectons et traitons environ **65 Téraoctets de données** dans deux **Data Centers**, mais c'est surtout **l'usage et le sens que l'on donne à ces données** qui sont **importants**



### QUELS ONT ÉTÉ LES PRÉREQUIS ET FACILITATEURS À LA MISE EN PLACE DU BIG DATA ?

**L'expertise data au sein de Voyages-sncf.com est ancienne mais à l'époque n'était pas formalisée sous le nom Big Data.** En effet, notre site, de par sa volumétrie, nécessite une grande technicité ainsi qu'un pilotage et un monitoring précis. Nous manipulons donc la donnée depuis très longtemps. **Voyages-sncf.com est un site avec une grande audience de plus de 12 millions de visiteurs uniques par mois tous supports confondus.** Nous avons donc la capacité de collecter un grand volume de données sur nos clients, ce qui est naturellement un prérequis au développement d'une véritable stratégie Big Data. Les données collectées le sont uniquement si nos clients l'acceptent, en sachant que Voyages-sncf.com ne vend pas ces données à des tiers.

Le Big Data nous permet d'adjoindre toutes les données de navigation à nos données clients, intégrées dans notre CRM. C'est un sujet que nous avons commencé à intégrer en 2013, sous la forme d'un POC (Proof of Concept), et que nous avons industrialisé à partir de 2014.

### COMMENT L'ÉQUIPE EST-ELLE COMPOSÉE ? COMMENT FAITES-VOUS LE LIEN ENTRE LA DATA ET LES MÉTIERS ?

Nous avons historiquement des équipes plutôt Data Mining qui géraient la manipulation de nos bases de données et avons intégré les compétences en Data Science en plusieurs étapes. Nous avons d'abord fait appel à des prestataires externes pour acquérir la connaissance et monter en compétences sur ces sujets puis nous avons recruté des Data Scientists en interne et nous formons maintenant des équipes data afin qu'elles puissent monter en compétences et maîtriser tous les outils de data mining et de Big Data. Nous disposons désormais d'une équipe de 15-20 Data Scientists, dédiée à la construction d'algorithmes utilisés en application sur les canaux marketing ou sur le site.

Nous sommes organisés en Feature teams au sein de Voyages-sncf.com. Ce sont des équipes qui regroupent toutes les compétences nécessaires à produire une nouvelle fonctionnalité et l'objectif du marketing est d'insuffler dans ces équipes la connaissance client aux collaborateurs qui vont développer le produit, faire du marketing relationnel et de l'acquisition. L'équipe data est au cœur de ces réflexions pour permettre aux collaborateurs de devenir le plus autonome possible. Dans le cas contraire, nous leur adjoignons un binôme data pour les aider sur leurs problématiques.

### QUELS TYPES DE DONNÉES RÉCOLTÉES ? QUEL VOLUME ?

Le volume de données récoltées est colossal. **Nous collectons et traitons environ 65 Téraoctets de données dans deux Data Centers, mais c'est surtout l'usage et le sens que l'on donne à ces données qui sont importants.** Nous combinons les données de transports, données clients, données de navigation et des données externes. **L'enjeu est de ne pas se noyer dans le volume de données mais bien d'arriver à dégager des usecases pertinents d'un point de vue business,** de donner du sens aux données pour la personnalisation et pour créer de la valeur pour nos clients.

Nous fonctionnons ainsi toujours sur la base d'une approche Test and Learn avec des logiques de déploiement partiel, de mesure avec des KPIs en amont et un suivi pour pouvoir en tirer les bonnes conclusions. Cela fait partie de notre approche agile du développement produit, un mode de fonctionnement essentiel chez Voyages-sncf.com. Nous avons donc une approche de développement progressif, pour éviter les effets tunnels et pouvoir apprendre et réajuster très vite.

Nous intégrons également une stratégie de co-construction tout au long de nos développements. Notre communauté d'utilisateurs peut tester les services et nous faire part de réactions et commentaires. Nous avons procédé ainsi avec la nouvelle page d'accueil ou encore avec le nouveau service « Mon Voyage ». **Les analyses data sont essentielles, mais il faut confronter et enrichir ces éléments quantitatifs avec la voix du client,** dans une démarche qualitative, pour donner de la chair et éclairer nos projets. La co-construction et les focus groups sont essentiels pour aider à la compréhension des besoins clients.

Les **analyses data** sont **essentiels**, mais il faut **confronter et enrichir ces éléments quantitatifs** avec **la voix du client**

## QUELS USAGES FAITES-VOUS DES DONNÉES ?

Nous avons trois typologies d'usages de la data. La première est technique et liée à la qualité de service et au pilotage du site web.

Nous utilisons ensuite les données autour des offres pour les exposer ou pour agréger les différents inventaires et calculer les itinéraires les plus pertinents. Nous proposons par exemple des trajets alternatifs, une nouvelle fonctionnalité qui permet au client de bénéficier, au-delà du parcours standard, d'un parcours à prix plus bas s'il a plus de temps pour voyager, en introduisant un transit ou un changement de gare.

La dernière typologie d'utilisation de la data est liée au marketing ; nous travaillons sur toutes les logiques de personnalisation, afin d'être plus pertinents dans les offres proposées à chaque client. La personnalisation est une des priorités de la feuille de route de Voyages-sncf.com. Cette année, la ligne directrice présentée par Franck Gervais, notre directeur général, a trait au smart tourisme, afin de fournir un accès fluide et sans couture aux offres proposées. Parmi les piliers d'une offre fluide, on retrouve la personnalisation ainsi que la data et la connaissance client exploitées pour l'ensemble des interactions client.

## QUELS ALGORITHMES POUR LES TRAITER ?

**Nous fonctionnons plutôt sur des algorithmes propriétaires, sauf lorsqu'il existe des standards adaptés sur le marché et un bénéfice à prendre des outils existants.** Sur les sujets cœurs de compétences, nous privilégions nos propres algorithmes.

## VOUS PARLEZ AU SEIN DE VOYAGES-SNCF.COM DE « SCÉNARIOS UTILISATEURS » : COMMENT LES DÉFINISSEZ-VOUS ET DANS QUEL BUT ?

Nous définissons des « scénarios utilisateurs » dans le but de cadrer et de relier nos projets à un vrai enjeu client. Nous nous assurons ainsi que chaque étape de la chaîne, des données collectées à la proposition de valeur mise en ligne sur le site, en passant par l'algorithme, est bien en cohérence avec nos objectifs clients.

Nous construisons nos « scénarios utilisateurs » en atelier de travail, ce qui nous permet de rassembler toutes les compétences autour d'un cas client : **Data Scientists, experts de l'UX, product owners, personnes des services études qui représentent la voix du client.**

Sur le sujet de la personnalisation, nous avons travaillé sur l'amélioration de notre home page ces derniers mois, avec pour objectif d'avoir une home page qui soit, à terme, différente pour chacun des utilisateurs. La spécificité de

Voyages-sncf.com est en effet d'avoir une audience très large, avec des utilisateurs plus ou moins experts. Certains de nos clients sont très aguerris à la réservation online et savent exactement ce qu'ils cherchent, mais nous avons également des familles qui font des recherches tarifaires plus compliquées, et enfin des personnes qui allaient auparavant en boutique, qui ont donc une maturité digitale moindre et un fort besoin de réassurance.

Cette variété de profils entraîne des besoins différents, d'où la nécessité de personnaliser la home page pour que chacun retrouve l'élément le plus important pour lui.

Nous avons également ajouté des services. Par exemple, quand un client a un voyage en cours, il peut voir remonter son voyage à venir sur sa home page personnalisée et avoir accès au nouvel espace « Mon Voyage ». Nous proposons également des contenus sur les activités

à la même typologie. **La personnalisation de nos campagnes, en marketing relationnel ou en display, a notamment permis une amélioration du taux de clic de plus de 30%.**

## QUELLES RECOMMANDATIONS POUR D'AUTRES ENTREPRISES ?

Il y a toujours des réticences lors du lancement de projets data, on se heurte régulièrement à des convictions très ancrées. Intégrer plus de data oblige à se défaire de certaines d'entre elles. Cela oblige également à structurer le travail de manière différente et à intégrer plus d'anticipation pour tirer véritablement profit des données. **On peut progresser sur la partie data, mais si l'on veut tirer toute la valeur de la data, cela nécessite d'aligner toute la chaîne autour des données.** Pour exemple, une fois



à faire lors de son voyage.

Nous avons créé Mon Voyage avec pour objectif d'accompagner le client, de l'achat de son billet jusqu'à son retour, dans une logique de personnalisation chrono responsive. Il s'agit de lui proposer des services utiles et complémentaires à chaque étape de son expérience de voyage, en fonction du timing et de l'état d'esprit correspondant, de sa sensibilité prix, de ses appétences etc.

## QU'EST-CE QUE L'UTILISATION DE LA DATA VOUS A APPORTÉ ?

Les chiffres sont véritablement encourageants, le domaine des destinations est celui dans lequel l'impact est assez visible. Quand nous souhaitons proposer une offre à nos clients, nos algorithmes de destination nous permettent de mettre en avant la meilleure destination à la fois sur la base des destinations passées de chaque client mais également sur la base des meilleures destinations de clients appartenant

que vous avez fait des ciblage très fins de clients grâce à la data, il faut avoir des créations (bannières, emailing) adaptées pour que cela soit bénéfique. C'est cet alignement qui peut prendre du temps.

**L'ancrage business est donc essentiel, il permet d'identifier les cas clients précisément, la valeur qu'on en attend et de définir tous les outils de mesure des bénéfices, avant même de mettre en place les outils data.**



ORNELLA  
**GODARD**

Neuroscientist - Feel Data Expert

**DATAKALAB**



NICOLAS  
**DELATTRE**

Feel Data Chief Strategist

**DATAKALAB**

# Feel Data : le Big Data émotionnel

**Ornella Godard** est diplômée de l'Institut de Psychologie de l'Université Paris Descartes. Elle est docteur en psychologie cognitive et est spécialiste en neurosciences et en neuropsychologie. Sa thèse de doctorat portait sur la reconnaissance des visages et des émotions faciales. Elle a été enseignant-chercheur pendant 8 ans et a fait un post doctorat sur le traitement des émotions faciales au cours du développement. Elle a utilisé des techniques variées comme l'électro-encéphalographie, des mesures comportementales et de l'eye-tracking (ou oculométrie). Après plusieurs années, elle s'est orientée vers l'étude du comportement hors laboratoire. Elle souhaitait utiliser ses connaissances neuroscientifiques au profit d'un aspect lié à la communication, au marketing et à la psychologie du consommateur. Elle a ainsi rejoint Perceptio Media et le projet de Datakalab.

### POURQUOI LA CRÉATION DE DATAKALAB ?

Datakalab est le fruit d'un rapprochement entre plusieurs sociétés, **Hémisphère Droit** fondée par Frank Tapiro, On Broadway production, société de production audiovisuelle intégrée et **Perceptio Media**, cabinet d'étude marketing spécialisé en neurosciences qui développe et collecte les Feel Data. Datakalab réunit des éthologues, des experts du digital, des ingénieurs informaticiens, des créatifs, des spécialistes de l'étude du comportement humain, des planneurs stratégiques... **L'objectif consistait à apporter de l'innovation marketing et à changer le paradigme dans une relation marque-consommateur plus du tout top down mais horizontale.**

Le premier constat concerne donc le marché. En effet, il est devenu plus difficile pour les marques d'innover et d'être créatives. On constate un véritable plafond de verre de la créativité et de l'innovation. Nous nous sommes également rendu compte que les sciences pouvaient apporter des réponses sur la compréhension des individus et surtout sur la façon dont on peut innover et créer de nouveaux produits et services en phase avec leurs émotions et leurs attentes.

### POURQUOI MESURER LES ÉMOTIONS ?

Les dernières avancées montrent à quel point l'émotion est prégnante et influence dans nos vies. Auparavant, on pensait que le consommateur était très rationnel et prenait des décisions basées sur le prix, sur le positionnement du produit mais l'on se rend compte que le consommateur est également très irrationnel, c'est un être humain.

Dès que l'on travaille en lien avec l'humain, les processus émotionnels sont partout même s'ils sont inconscients, ils influencent nos comportements, nous font réagir face à un stimulus extérieur. Ce sont des processus profondément ancrés et façonnés tout au long de l'évolution. Les sociétés qui développent des techniques d'analyse du comportement du consommateur

**Nicolas Delattre** a travaillé plus de 15 dans le business development au sein de grandes sociétés. Nicolas et son associé Alexandre Le Texier, avaient toujours eu envie de mener une aventure entrepreneuriale. Après une rencontre déterminante d'Alexandre avec un directeur de laboratoire de Lille, spécialisé dans les sciences cognitives, ils ont conjointement décidé de lancer Perceptio Media en janvier 2013, convaincus de la capacité des sciences à apporter au marketing. Ils sont en cours de rapprochement avec Datakalab.



se basent sur le déclaratif, le plus souvent sur des focus groupes plutôt que sur des entretiens individuels. Tous les grands instituts d'études des marques savent qu'il s'agit de méthodes limitées car elles comportent de nombreux biais sociaux et d'expression.

Malgré l'existence d'outils neuro-scientifiques utilisés en laboratoire depuis de nombreuses années pour comprendre le lien entre cerveau et comportement, ce n'est que récemment que ces outils sortent des laboratoires pour permettre l'étude comportementale en situation réelle et la collecte de données objectives et modélisables. **C'est un vrai challenge mais nous sommes persuadés que la recherche et le marketing peuvent se booster mutuellement.**

Les bénéfices de l'émotion sont nombreux parmi lesquels les réflexes de survie, l'encodage mémoriel, l'influence sur la perception du monde et l'action notamment dans des situations dangereuses mais pas seulement. L'émotion est une action en soit et elle conditionne en retour notre comportement.

L'émotion est multi-componentielle et se traduit par des aspects physiologiques comme la conductance électrodermale, l'accélération du cœur, la dilatation de la pupille, des aspects comportementaux comme les expressions faciales, l'intonation vocale, la posture ainsi que par des composantes subjectives et cognitives quasiment impossibles à modéliser de manière informatique. **Cet aspect cognitif de l'expérience émotionnelle vécue peut se traduire verbalement d'où la nécessité de coupler des indices physiologiques objectifs aux déclarations verbales des consommateurs, qui sont également porteuses d'informations.**

### LES FEEL DATA, QU'EST-CE ?

Ce sont les données émotionnelles des individus, d'un public, d'un consommateur qui sont collectées au moment même où l'individu vit une expérience de contenu audio-visuel comme une publicité, un film mais également des expériences de produits, de services...

Nous utilisons un **dispositif fixe** de mesure de ce que perçoit (mouvements oculaires et dilatation pupillaire) et ressent une personne (intensité émotionnelle via la mesure de la réponse électrodermale). Ce dispositif combine donc plusieurs outils sophistiqués utilisés dans la recherche fondamentale ou médicale, permettant de collecter en temps réel les données physiologiques qui traduisent **la perception visuelle, la charge cognitive et l'intensité émotionnelle d'un individu face à tout type de contenu audiovisuel.**

1) **L'enregistrement des mouvements oculaires (oculométrie ou « eye-tracking »)** permet d'analyser la perception visuelle et d'enregistrer **la position de l'œil sur l'écran** toutes les millisecondes, le temps de fixation oculaire et **les saccades** d'un point A à un point B quand on explore l'environnement.

2) Nous enregistrons également **le phénomène de dilatation pupillaire** qui est une réaction physiologique qui se produit quand on a plus ou moins de lumière qui rentre dans la rétine, mais qui reflète également la **charge cognitive** engagée dans une tâche. Nous avons ainsi développé **un algorithme permettant de neutraliser l'influence de la lumière naturelle pour ne retenir que l'effet de la charge attentionnelle.**

3) **La réponse électrodermale**, liée à la micro-sudation de la peau, est un indice physiologique reflétant l'intensité émotionnelle vécue par l'individu en temps réel. La peau est composée



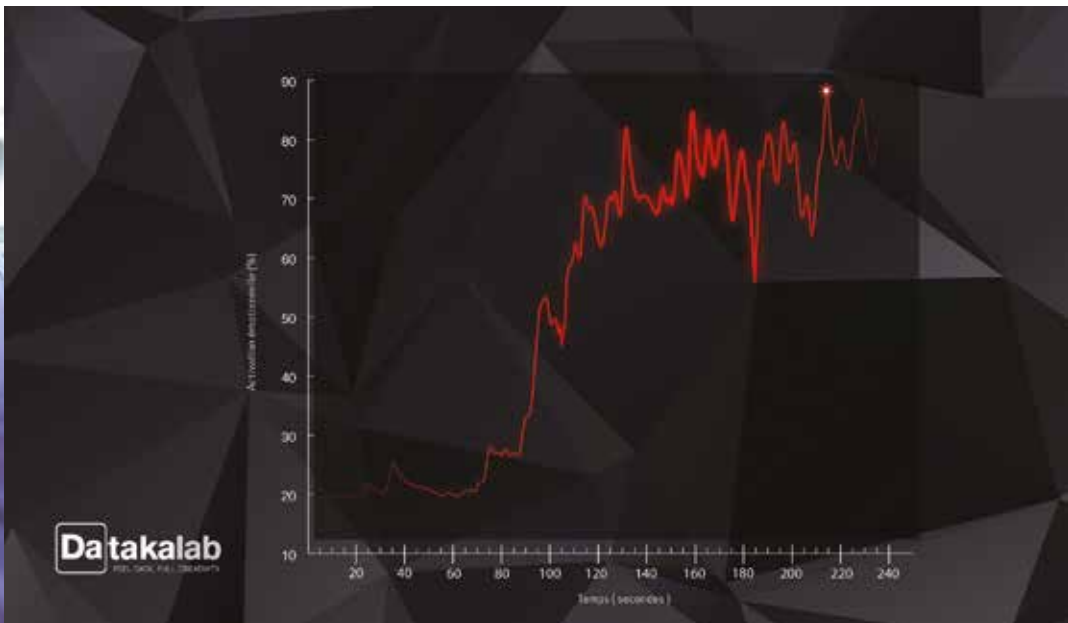
de micro canaux qui se remplissent de micro gouttelettes de sueur, en fonction de l'intensité émotionnelle. Il s'agit d'un faible courant électrique qui passe sur la surface de la peau, qui nous permet de mesurer l'effet de conductance, l'intensité émotionnelle s'exprimant au travers de la rapidité de la conductance. Le courant passe et le capteur enregistre les variations du courant de la peau à des endroits précis de la main.

Nous avons fait le choix de prendre les meilleures technologies du monde de la science telles que les technologies Eye Link, nous n'en sommes donc pas propriétaires. En revanche,

sonne, où qu'elle se trouve et quelle que soit l'expérience vécue. Le bracelet permet également d'enregistrer des **données cardiaques**. La technologie embarquée ouvre des champs très larges. On sort vraiment de l'écran, on peut mesurer l'émotion d'une personne, qui va manipuler une application sur un téléphone, qui va vivre une expérience en magasin, lors d'un spectacle, ou dans un musée pour retracer le chemin émotionnel d'une exposition artistique, dans des parcs d'attraction. Cela peut servir de nombreux domaines, par exemple sur des mécaniques émotionnelles liées à des storyboard, à du design, de l'architecture, à la santé, au bien-être ou encore à la relaxation.

### COMMENT OBTENIR L'ACCORD DES PERSONNES POUR TESTER LES FEEL DATA ?

Dans les études marketing, même si les participants peuvent être rémunérés, ils signent un consentement libre et éclairé. Les données sont conservées sur un temps limité avec un objectif précis. Nous suivons une charte déontologique et toutes les données sont anonymisées comme pour toute étude marketing.



notre expertise réside dans le développement des logiciels et algorithmes pour traiter et collecter ces données. En l'occurrence notre système va synchroniser ces données, nous permettant ainsi de faire des analyses croisées et d'avoir une vue globale de ce qu'un contenu a provoqué et pour quelles raisons. Pour les marques, comprendre les raisons d'un succès ou d'un échec dans un changement de communication est primordial pour continuer à innover et avoir un temps d'avance. Ces techniques sont utilisées pour tester des contenus audiovisuels, des spots tv, du brand content, du print, des bandes annonces de films, la home page d'un site web, le packaging d'un produit, une œuvre d'art ou même une architecture. Les contenus peuvent être fixes ou animés.

La collecte non verbale est une innovation mais nous la couplons toujours à des techniques verbales d'entretien, d'autant plus pertinentes que l'intervieweur a les résultats de ses Feel Data sous les yeux. Cela enrichit la capacité d'investigation et donne une perspective plus facile pour obtenir des informations spontanément. Il y a une notion d'empathie qui libère naturellement la parole, nous ne sommes ni dans le jugement ni dans l'intrusif. **Déclaratif et non déclaratif se combinent sans heurt, en effet les capteurs n'auront jamais réponse à tout.** Nous utilisons également une solution embarquée, sous la forme d'un bracelet qui ressemble à une montre et qui mesure cette fois-ci uniquement l'intensité émotionnelle d'une per-

### LES FEEL DATA SONT-ELLES CROISÉES À D'AUTRES DONNÉES ?

Au sein de Datalab, nous utilisons différentes données pour favoriser la créativité et l'innovation, les **Feel Data** principalement mais également **les données des marques, leur ADN et leurs données CRM.**

**Les data émotionnelles représentent le nouveau carburant de l'industrie de la data.**

Le « Graal » serait de pouvoir corréliser les Feel data aux données du Big Data, d'enrichir ces données et développer des modèles prédictifs. On parle beaucoup d'insights, l'idée pour en générer est de multiplier les **sources d'informations pertinentes** qui vont pouvoir être transformées pour générer de nouvelles informations : c'est l'essence même du Big Data.

### QUELS SONT LES SECTEURS DANS LESQUELS LES FEEL DATA PEUVENT ÊTRE APPLIQUÉES ?

Les Feel Data ont plusieurs champs d'application : **LE BRANDING**, la communication, les campagnes publicitaires

**L'EXPÉRIENTIEL**, pour les émotions ressenties par le shopper en point de vente ou dans les flagships, mais également sur les produits et services eux-mêmes

**LES TESTS D'APPLICATIONS DIGITALES ET LES TESTS D'INTERACTION HOMME-MACHINE**

Dans ce 3ème champ d'application, nous utilisons alors une combinaison d'outils : le bracelet et l'oculomètre, spécifique aux smartphones ou tablettes, développé par la société TOBII.

Un grand nombre d'entreprises développe des



**Déclaratif et non déclaratif se combinent sans heurt, en effet les capteurs n'auront jamais réponse à tout**

applications et ne mesure que des indices très rationnels et chiffrés, comme le nombre de pages vues, le temps passé par page... Ces indices permettent de faire un constat sur le comportement final observé mais non d'expliquer les actions intermédiaires ni les raisons de ces actions. D'autant qu'une application se doit d'être rapide, bien designée car son rôle est de rendre service et de gagner du temps. Ainsi, les mesures du parcours oculaire et des émotions au cours de l'utilisation des applications fournissent des informations précieuses sur ce que procure l'expérience digitale et comment façonner une application qui génère telle ou telle action chez l'utilisateur.

### 1<sup>ER</sup> CAS : LE CAS D'APPLICATION PMU : « VIVEZ VOS ÉMOTIONS À 100% »

Nous avons travaillé sur une campagne publicitaire du PMU dans une période de refonte de l'identité visuelle autour de l'émotion en 2015. Leur nouvelle baseline était « Vivez vos émotions à 100% ».

Lors de la première vague de communication sur leur nouvelle campagne, composée de plusieurs spots tv et d'affichages print, les résultats n'étaient pas à la hauteur de leurs espérances. Nous avons donc effectué un **diagnostic pour évaluer la puissance émotionnelle de leur spot**. Nous avons recruté 20 personnes qui correspondaient à la cible de la marque et leur avons fait visualiser les différentes publicités suivi d'un entretien individuel. Nous avons ainsi pu fournir des analyses et rapports assortis de recommandations qui ont permis de modifier certains éléments de la campagne. Le spot a été retravaillé, la musique a été changée, la mécanique de storytelling ainsi que différents éléments visuels ont été modifiés, afin de libérer le potentiel et l'impact émotionnels du spot.

L'idéal est naturellement d'intervenir en amont, au niveau de la stratégie, pour proposer des solutions et orienter vers un choix plus pertinent. Les Feel Data apportent des données objectives pour aider les agences à vendre un projet de campagne ou leur permettre de minimiser les allers-retours. Dans la publicité, le but n'est pas d'en faire plus, car les consommateurs ressentent une certaine saturation publicitaire mais de faire des communications mieux ciblées, plus engageantes et plus émotionnelles, afin d'activer plus efficacement les consommateurs.

Les **data émotionnelles** représentent le **nouveau carburant de l'industrie de la data**

### 2<sup>E</sup> CAS : LES FEEL DATA LORS DES ÉLECTIONS PRÉSIDENTIELLES : UN CAPTEUR DE VÉRITÉ ?

C'est un domaine d'application possible en tant qu'électeur, c'est une question de société. Quand on parle d'émotion par rapport à la politique, l'idée est de dire que l'émotion est générée par l'empathie ou l'antipathie qu'on peut avoir pour une personne. **Ce qui va générer de l'émotion auprès des gens, ce ne sont pas les discours exaltés mais la sincérité d'un politique**. Si l'on pouvait mesurer les émotions, cela libérerait la parole des politiques. **Utiliser les Feel Data peut être un moyen de réinventer le dialogue entre les concitoyens et leurs représentants dans une période de crise politique**. Les Feel data sous le prisme de l'émotion permettent plus de sincérité. L'élection présidentielle c'est l'élection d'un homme, d'une personne et cela inclue un côté irrationnel, viscéral et physique. Le lien émotionnel doit être valorisé et doit aider à prendre une décision. Nous n'envisageons pas d'utiliser les Feel Data pour renforcer l'effet « show, » comme « l'effet Trump », mais bien de l'utiliser dans une optique de sincérité, vérité, bienveillance et de redonner du lien. Un concept d'émissions dans lesquelles les émotions sont mesurées lors d'un panel, est tout à fait envisageable.





BALA  
**GOPALAKRISHNAN**

VP Innovation, Analytics  
and Technology Architecture

**THE WEATHER NETWORK / PELMOREX**



**Big Data:**  
more than just  
a **fair-weather**  
**friend**

**Bala Gopalakrishnan** has a background in technology and holds a Master's in Computer Engineering from the University of Alabama in USA. Prior to joining Pelmorex, he worked in the technology and software development field. His first position was with Johnson Controls, a Fortune 100 company in developing hardware and software for computer-based controls industrial/building systems. He then joined a start-up, which grew into a larger organization, called Eutech Cybernetics, working in the IoT/Smart Cities technology integration sphere. Having completed his MBA in Canada, he was keen to work at the cross intersection of technology and business, inciting him to join Pelmorex, which is an innovative leader in bringing new products to consumers combining technology, science and creativity. He has been with the company for approximately 10 years in multiple roles, working on content, product and technology.

**WHAT IS PELMOREX'S CORE BUSINESS?**

Pelmorex is one of the largest data and media organizations in Canada dedicated to the weather. **It has large number of consumers - about 46 million users globally.** From a content perspective, a lot of data is generated. Also, the users generate audience and behavioural data through interactions. Choosing to work for the firm gave me a good opportunity to further develop my expertise and learn about Big Data. It is a company where the data is very fertile, and as the Big Data revolution grows, we are only natural to capitalize on that moving forward.

The Big Data side of our business has become much more active over the last 2 years. When we looked at big players such as Google or Facebook, **we realized that one of the biggest assets we had going forward was data.** We produce a lot of content: weather forecasts and micro climate forecasts are generated every few minutes to every few hours. We have an editorial news team that write about the weather and climate change. The company has 24/7 television channel providing people with information about the weather and generating lots of video content and also live video streams, interactive apps on internet for over the top video devices like Apple TV, Android, TV, Roku etc. We combine all that with the potential mix of millions of users, who open our website and mobile applications numerous times a

day, generating their own local weather photographs, videos, submitting information whenever there is weather activity or any news to report. This all represents a massive data asset that we can use in many different ways to both engage customers and create useful products for them.

### WHAT IS YOUR ROLE AS VP INNOVATION, ANALYTICS AND TECHNOLOGY ARCHITECTURE?

My current role reports to the organization's CIO and has three main functions.

The first is related to the amount of data we collect and generate. In order to process all the input raw data and make end consumer products available on multiple distribution platforms such as televisions, large multi-language websites serving several different countries, mobile apps on iOS and Android, iPads, Tablets, smart TV and OTT devices, we need to have the right Architecture in place. I lead this area to ensure **we have the right technology systems in place, capable of supporting the various media platforms.**

Then there's thinking about how to mine this data effectively in order to gain insights we can then convert into new, useful projects. Our product is free to end-users, everything is monetized through advertising. We consequently want to use data in two ways - **to derive useful consumer insights, so they find the content really engaging and to target them with the right advertising.** Both perspectives rely on analytics.

Any company needs to innovate to stay current and useful to the consumer in today's changing world, and to us innovation comes from combining the data we have with technological advancements. Today, people even access weather reports via smart watches, so it is vital we constantly innovate to remain leaders on different platforms. Computers are now more powerful, there are new ways of processing meteorological information, new advances have been made in the meteorological science field. **Innovation is a combination of all this progress in technology, data and science.** I coordinate or lead projects that bring those aspects together. My team combines lots of these advancements to create new products and services, which they then prototype, test in the marketplace and present to the management team and key stakeholders. That then makes it easier for us to decide each year which new consumer initiatives we are going to launch.

### TO WHOM DO YOU REPORT WITHIN THE ORGANIZATION?

My current role involves working with four different types of people: data people, such as data scientists or data analysts; front and back end developers; a meteorological science team who are largely weather scientists studying algorithms; and finally my clients, who are my business partners in the company: Product leaders on consumer side and Advertising sales folks who monetize and generate revenue using the insights we have on the advertising side.

My company has realized that data represents a massive opportunity, so I am going to move to a new role that completely focuses on data - Managing Director of Data. That means I've progressed from combining multiple roles, to a highly specific senior role focusing on nothing but data. This role reports to the CEO. The CEA

and board are definitely convinced of the importance of data - the fact we have so much data on what is happening in the marketplace was enough to convince everyone.

We still have to deliver a lot of positive results to prove we have made the right choice. We have already seen some promising results and early successes, but we are still 30-40% from being a totally data-driven organization.

### CAN YOU TELL US ABOUT HOW DATA IS USED AT PELMOREX? WHAT KIND OF DATA DO YOU COLLECT?

We collect and produce **weather data**, primarily weather forecasting data such as what the weather will be like in the next few hours, in the next fortnight or over the next few months. We have specialized products like minute by minute forecast of rain and snow in next hours, long range weather outlook on calendar and seasonal forecasts. In itself, weather is a huge data set including observation stations, weather camera images, hourly forecast, radar, satellite imagery coming in from hundreds of thousands of locations around the world. 180 million forecasts are generated by our Forecasting engine for every kilometre in Canada, for instance. And we also draw on former weather patterns and climatological data, using 50 years of historical weather data for different places around Canada.

We get around **1 billion location records every month** from different users

We also have **articles and videos** explaining both the in the moment weather data but also discussing topical issues such as climate change. One of our biggest assets is **user interaction**. For instance, you could be looking at the fore-

cast today in Paris, but you may be planning to fly to Toronto in 2 days' time. We might notice when you use our app that you are starting to look at the weather in Toronto, allowing us to infer that you may be planning a trip. We can consequently bucket you into our business traveller segment for Toronto, allowing us to advise our partners to push relevant advertising deals on business hotels. From a content perspective, meanwhile, we might show you videos or articles related to Toronto. That is what we call behavioural audience data.

Another aspect relies on the fact that the **weather is always consulted in a location sensitive context, related to where people are or where they are going.** We are one of the services that asks people to turn on their GPS in order to use our application, and the acceptance rate is somewhere in the region of 75% to 80%, which is really high. This alone is a data set, as we **get around 1 billion location records every month from different users**, which we then use to create new consumer products.

In Toronto, there is a place called Toronto Island, which gets very busy in the summer. Based on location data, we can actually analyse the times at which traffic will be busy on the island and then proactively use that information to send push notifications or alerts to people living close to the island or who go there on a regular basis, telling them it's the perfect weather to go there, but warning them to set off early as it's likely to be crowded.

Our company philosophy is that **"people use weather information to make decisions"**, either when planning something or for safety reasons, and we can help by combining the audience, behavioural and location information we get and the unique content we have.

We have a lot of unique datasets, such as the surface and conditions of highways and roads, and the places where lightning strikes. We even have data related to weather and illness, such as where pollen is found in the summer, that we can use to help prevent allergies.

### WHAT KIND OF ARCHITECTURE DO YOU USE?

**Our architecture is described as a hybrid.** A lot of our audience data is collected using Google tools from Google Analytics, and we use Google





Big Query, which is a Google Big Data Stack. For our users' behavioural data, we use Microsoft Azure, HD Insights and Stream Analytics to capture real-time information. For predictive modelling and to detect correlations between the weather and other parameters we use a combination of Amazon Web services, Spark and some in-house clusters.

**We use all the stacks, and our technological decisions are based on what is cost effective and what is the best option in terms of cost of data transport for a specific use case.** For example, as our audience data is generated through Google Analytics, it is easier for us to transfer that data into Google Big Query, and then carry out our analyses there and connect the dots.

It saves us both money and time because the two systems are interconnected and we can rapidly deliver meaningful insights.

On the weather side of things, we run our meteorological algorithm to get better forecasts using in-house clusters, because **both the algorithm and the science are highly proprietary.** We use different types of models - American, Canadian and European - , so our meteorologists have built a **model that blends up to 30 different models.** Usually, most of them offer information covering a surface area of 15 to 50 kilometres. Our algorithm takes things down to a much finer level, blending multiple models and applies proprietary algorithms to produce weather forecasts for a specific postcode or neighbourhood covering up to one kilometre. For instance, I live close to the lake, so the general model may not be particularly relevant, as where I live is cooler than the surrounding area. Our algorithm takes that into account and produces a microclimate forecast, specific to the surrounding 1km.

We also create a system of virtual observations. There are real observation stations around the world with sensors in them, which detect the temperature, humidity etc.... but they are limited in number. To solve this problem and create more observation points, we use the forecast information for the next 15 minutes combined with radar information, which is more granular. **Virtual observation is very accurate as it mainly relies on computer engineering processes to extrapolate data based on various data sources.**

### *HOW SUCCESSFUL IS YOUR WEATHER PREDICTIVE ANALYSIS?*

Prediction science was around well before Big Data came into play, but Big Data brought three new things:

- **much more input data;**
- **much more processing,** meaning we can blend many more models;
- **much more granularity:** in the past you would get the weather forecast for a city, but today a city can itself have several different microclimates.

Recently, Big Data has also allowed us to create a new algorithm called "when is it going to rain", but down to minute by minute prediction from 30 minutes to 6 hours' timeslots. That type of time precision for a specific microclimate is one of the areas in which Big Data has progressed, enabling us to create the precipitation start-stop algorithm. It is between **70 to 100% accurate, which is pretty impressive.**

### *HOW DO YOU CONNECT ALL THIS DATA TO THE USERS OF THE APPLICATIONS?*

Our company's motto is: "**Weather is there to serve consumers' planning and safety needs**". We constantly strive to understand consumers, what their daily planning and safety needs are. We also try to identify which needs are not yet fully met. **Our data teams examine these challenges and try to figure out what can be done to better address the various needs, attempting to find the right user experience by combining data, UX, product and technology.** The precipitation start-stop is one of the problems that we identified as an important consumer need.

There are a lot of new products that can be generated by data itself. We have access to huge amounts of consumer data relating to both behaviour and location, and this data fuels new product concepts. Our typical prototyping cycle is about 3 weeks long; we try to fail quickly. Taking the product from the drawing board to the market - or not - typically takes around a couple of months to few months. For instance, the weather affects the mood of our users, so we are thinking of developing an application that allows people to give feedback on how they feel about the weather, asking users what their mood is in order to create a metric such as the mood of the day or the mood of the hour. This would be a totally crowdsourced product driven by its users.

### *WHAT IS THE ROADMAP OF INNOVATION AT PELMOREX?*

**In terms of Connected Things, it has already been multiplied by 100, and we expect the amount of data flowing in to be 100,000 times bigger.** The more information we have, the better the predictions will be. Our challenge is to cope with all this information coming in and to manage it without having huge capital outlays. We also need to filter the noise - there is a lot of information coming in but lots of it can be misleading.

We can create very useful and granular micro climate forecasts, but at the end of the day that needs to translate into useful insights for each person's specific context. How do we go about understanding the context where each person is standing? Understanding the types of decision, they are trying to make? For instance, if you are in an airport, looking at the weather forecast, you probably need to make a different decision than you would if you were standing on a soccer field with your children. The second challenge is consequently **to provide consumers with the right product, using the location signal, user behaviour data and all the information available about them to understand and use the context, thereby delivering products that reflect each user's context.**

Finally, weather-related decision making for enterprises is another area that is opening up to us. Every business is affected by the weather. Business supply chains, sales activities and even product pricing can all change based on the weather conditions. This is something we plan to develop in the future.

In terms of **Connected Things**, it has already been **multiplied by 100**, and we expect the amount of data flowing in to be **100,000 times bigger**



## RAPHAËL CHERRIER

Founder and CEO

**QUCIT**



**Raphaël Cherrier** a un parcours scientifique, physique et mathématique. Il a fait Normale Sup et a obtenu une thèse en physique théorique. Il a été Maître de conférences en physique également. Il aborde le monde en essayant de trouver les principes qui régissent la dynamique de la nature. C'est ce qu'il a souhaité faire en créant Qucit : mathématiser la ville pour prédire les comportements humains, la mobilité mais également le ressenti des citoyens par rapport à leur environnement proche. Nous avons souhaité interviewer Raphaël Cherrier pour connaître sa vision de l'impact potentiel de la Data sur la ville intelligente.

# Mathématiques et prédictif au service de villes durables et intelligentes

## QU'EST-CE QUI VOUS A POUSSÉ À CRÉER QUCIT ?

J'ai souhaité créer Qucit pour deux raisons principales : d'une part, la volonté de **mathématiser le monde** et de comprendre les principes fondamentaux qui régissent l'univers ; d'autre part la protection de l'environnement et notamment l'envie de participer à la création de villes plus durables, plus écologiques avec moins de gaspillage des ressources naturelles. Qucit est l'intersection de ces deux aspirations et correspond à une volonté d'avoir un impact sur la société, ce que la Recherche Fondamentale ne permet pas forcément en termes de moyens d'action. Au sein de Qucit, je pense réellement avoir la possibilité d'agir, de continuer à faire de la recherche poussée tout en ayant un impact direct sur la société.

On dispose aujourd'hui d'un volume de données qui nous permet d'étudier avec précision l'environnement extérieur. **La ville est entièrement numérisée et les quantités de datas produites dans les villes vont continuer à doubler tous les 2 ans**, suivant la Loi de Moore. Au fur et à mesure, nous allons obtenir une vision bien plus détaillée de la ville qui pourra être soumise à l'analyse mathématique prédictive. En parallèle, nous avons une vision bien plus précise de ce que font les êtres humains équipés d'un téléphone portable qui capte leurs déplacements ou leurs modes de transport.

La ville est entièrement numérisée et les quantités de datas produites dans les villes vont continuer à doubler tous les 2 ans

## COMMENT AVEZ-VOUS CONSTITUÉ VOTRE ÉQUIPE ? AVEZ-VOUS DES DIFFICULTÉS À RECRUTER ?

Notre équipe est essentiellement scientifique et technique, nous avons mené beaucoup de projets de R&D ces deux dernières années, cela se reflète naturellement au niveau des talents recrutés. Au sein de Qucit nous sommes : 3 docteurs en maths, informatique et physique théorique, des Ingénieurs de Centrale Paris, de l'ENSEIRB... qui constituent un pôle de 4 Data Scientists, un développeur front end, application mobile et Dashboard, une graphiste, un CTO qui s'occupe également de l'infrastructure Cloud et un directeur commercial. Nous sommes assez stricts au niveau de nos recrutements et nos exigences ne font qu'augmenter. Au niveau commercial, notre objectif est de gagner de l'argent tout en rendant les villes plus durables et plus agréables : **il s'agit d'un véritable challenge notamment en termes de business model**. Au niveau intellectuel et technique, travailler sur les sujets liés à la ville intelligente s'avère très satisfaisant. Ainsi **d'excellents profils nous ont rejoints car ils sont convaincus que cela a du sens**.

## QUELLE EST VOTRE APPROCHE ET VISION DU BIG DATA ?

**Tout d'abord, le Big Data est pour nous un outil. Je préfère d'ailleurs parler de Smart Data pour ce que nous faisons.** Google ou Facebook font face à de véritables problématiques de Big Data dans la mesure où aucun outil existant ne répond à leurs problématiques de gestion des données de leurs utilisateurs et d'indexation de l'ensemble de l'information disponible. Au sein de Qucit, nous apprenons à utiliser parfaitement les meilleurs outils et logiciels open source existants, par exemple Tensor Flow que Google a récemment mis en open source. Notre cœur de métier est de créer des modèles de Machine Learning et une infrastructure logicielle adaptée pour répondre aux problématiques des villes.



Nous développons la capacité à traiter toutes les données de la ville, afin de produire un modèle numérique de la ville. Nos algorithmes sont capables de traiter tous types de données en entrée et de prédire tout type de quantité en sortie. Nos algorithmes utilisent par exemple les données Météo, de circulation, d'occupation des stations de Vélib, des données Open Street Map, et toutes celles que l'on peut trouver en Open Data, que ce soit la position des bancs et des arbres ou des données produites par des devices. Ces données sont toutes intégrées et traitées automatiquement par nos algorithmes : elles permettent de faire des prédictions telles que le taux de remplissage d'une station de Vélo en libre-service, le temps pour trouver une place de stationnement en voirie, les déplacements en transport en commun et même le confort ou le stress ressenti par les piétons sur la place de la Nation.

## LES MODÈLES PRÉDICTIFS AU SERVICE DU TACTICAL URBANISM

Nous faisons actuellement des expériences sur la place de la Nation qui consistent à modéliser les sentiments humains. Toutes les données sont identiques en entrée en revanche de l'autre côté de la chaîne, les capteurs sont les êtres humains. **Nous mesurons les sentiments afin de déterminer si les individus se sentent stressés, en sécurité, désorientés à tel endroit et à telle heure.** Nous allons jusqu'à leur demander s'ils trouvent que leur environnement proche est propre ou beau. Les réponses à ces questions nous permettent de **calibrer des mo-**

## dèles d'Intelligence Artificielle qui prédisent et expliquent les facteurs qui font qu'un lieu public va être perçu comme beau par les gens.

Sur la place de la Nation, nos modèles nous ont par exemple permis de trouver les endroits qui donnent le meilleur point de vue sur la place et la statue centrale.

Pour étalonner nos modèles, il nous a donc fallu organiser des sondages : des étudiants de l'Ecole d'Ingénieur de la Ville de Paris ont donc interrogé 1300 personnes, pendant 4 jours, sur la place de la Nation, sur la base d'un questionnaire d'une quinzaine de questions. Cela peut paraître beaucoup mais c'est un échantillon très faible pour calibrer un modèle complexe avec autant de variables contextuelles.

Le concept derrière cela est physique. En physique, il y a le champ électro magnétique et ce champ est créé par les charges électriques. Un électron va créer un champ électrique et quand il se déplace il crée également un champ magnétique. La physique décrit les lois qui font le lien entre les sources et le champ produit. Chez Qucit, nous avons souhaité appliquer le même concept. On va ainsi avoir un champ de sécurité, un champ de stress et de confort. **Nos modèles nous permettent de déterminer quelles sont les sources du confort ou du stress.** Concrètement, cela peut venir du fait qu'il y ait à la fois des bancs et des arbres à proximité pour permettre de s'asseoir et de lire. Les sources négatives apparaissent également comme la circulation, la pollution, le bruit. Nous allons ainsi non seulement trouver les sources pertinentes pour prédire l'évolution de chaque champ mais également trouver les lois quantitatives de cette évolution. Ainsi le contexte urbain est pris en compte dans les modèles, mais aussi des sources internes à chaque personne qui sont fonction de l'âge, du genre, de la CSP, du lieu de résidence, etc. Pour exemple, une personne qui habite juste à côté de la place de la Nation va naturellement être moins désorientée que quelqu'un qui habite à l'autre bout de Paris et cela ressort vraiment du modèle. Mais de nombreuses autres sources plus subtiles ressortent également du modèle.

**Nous mettons notre capacité de modélisation au service du réaménagement de la place, pour faire du Tactical Urbanism.** Nous ne voulons plus qu'une place soit construite et que les défauts d'aménagements soient constatés à posteriori, mais rendre possible des modes de fonctionnement itératifs sur le terrain.

L'objectif de l'expérimentation sur cette place, est de tester les aménagements avec des barrières provisoires ; de compter, à l'aide de multiples capteurs : les flux piétons, vélos et automobiles ; de vérifier que cela ne crée pas de nouveaux problèmes et de mesurer en parallèle les sentiments des piétons. Compter les piétons sur un trottoir est plutôt aisé. En revanche, identifier si les piétons se baladent

**nombreuses subtilités.** Il pourrait être nécessaire de faire des modèles plus avancés, pour apprendre par exemple au modèle qu'il est en train d'analyser une place, avec des indicateurs tels que les routes qui tournent en rond, les terre-pleins centraux, les arbres et fontaines etc. C'est une des pistes d'amélioration de notre modèle.

## LES APIS AU CŒUR DU BUSINESS MODEL

L'application BikePredict est aujourd'hui disponible dans quasiment toutes les villes de France et plus généralement d'Europe qui proposent des vélos en libre-service et nous disposons de datas pour le faire dans plus de 500 villes du Monde. L'intégration du prédictif permet de garantir à l'usager la disponibilité d'un emplacement pour pouvoir garer son Vélib'. Par ailleurs, notre API peut servir à un opérateur qui veut offrir des fonctionnalités prédictives à ses usagers. Elle peut également être intégrée dans des calculateurs d'itinérance multi modaux.

L'idée est de proposer un trajet multimodal optimisé, par exemple entre La Défense et République. Afin de pouvoir proposer de prendre



ou vont travailler recèle plus de subtilité. Statistiquement, nous allons être capables de trouver des patterns comme par exemple une parcelle de trottoir utilisée par des gens qui commutent ou qui ont une activité de loisir.

Pour cette expérimentation, la place de la Nation a été équipée de capteurs, par Cisco notamment, mais également par des startups qui ont fourni les capteurs qui nous permettent de suivre la modification de la dynamique sur la place. Notre application gratuite 'Confort Paris' permet également aux parisiens de continuer à participer aux sondages pendant la phase de réaménagement. Nous allons ainsi pouvoir observer l'évolution du ressenti des piétons et autres usagers de la place et déterminer si les réaménagements conduisent à une amélioration. Les modèles étalonnés sur la place de la Nation sont répliquables. Ils nous permettent par exemple d'établir une carte du confort, du stress également sur la place d'Italie sans refaire de nouveaux sondages, même si une nouvelle campagne de sondage donnerait une meilleure précision. Un point intéressant, par exemple, c'est qu'il y a un supermarché sur la Place d'Italie alors qu'il n'y en a pas Place de la Nation. Mais notre modèle, même s'il ne sait pas qu'il s'agit d'un supermarché, sait prédire à partir du reste des informations contextuelles que cela ne va pas être beau et agréable devant le supermarché.

Sur ce projet nous n'utilisons pas de Deep Learning car notre échantillon de 1300 répondants ne serait pas suffisant pour calibrer un modèle de ce type. **Nous utilisons des modèles de Machine Learning standards que nous maîtrisons parfaitement et qui révèlent déjà de**

Nous mettons notre  
**capacité de modélisation**  
au service du réaménagement  
de la place, pour faire du  
**Tactical Urbanism**

le RER A jusqu'à Châtelet puis le Vélib', il faut être sûr que l'utilisateur dispose d'un Vélib' à Châtelet et d'une place de Vélib' pour se garer à République. Il est donc impératif de faire du prédictif. Cela permet d'intégrer une expérience utilisateur de qualité lors de l'utilisation d'un vélo en libre-service avec d'autres modes de transport en commun. Notre API est également un outil permettant d'optimiser l'équilibrage des stations par l'exploitant. D'une part, cela permet aux opérateurs d'optimiser la logistique de déplacement des vélib's ; d'autre part les utilisateurs en suivant les recommandations de BikePredict, participent également de manière collaborative à l'effort de rééquilibrage du réseau.

L'intérêt des APIs est d'enrichir les applications existantes. Avec des APIs, nos clients ne sont pas obligés de développer une nouvelle application, la brique peut être intégrée à celle existante pour permettre de conserver leur base utilisateur.

**Pour résumer notre stratégie, nous souhaitons donc résoudre les problèmes du vélo en libre-service, puis celui de la mobilité et enfin participer à rendre les villes plus durables et plus agréables.**





ALLAA R.  
**HILAL**

Innovation Lead

**INTELLIGENT MECHATRONIC SYSTEMS (IMS)**



# Data for **smarter, safer and greener vehicles**

**Allaa R. Hilal** has over 10 years' experience in Data Intelligence, IoT, and Intelligent Systems design. She holds a Doctorate degree in Electrical and Computer Engineering from the University of Waterloo and specializes in Pattern Analysis and Machine Intelligence. Dr. Hilal has been granted numerous prestigious awards and has also been recognized as one of the most influential women in the M2M and IoT field by Connected World Magazine.

She has always been an ardent supporter of women in science technology and engineering, and for the past 9 years has been an active member of the Women in Engineering group at the University of Waterloo. She champions the Women in Technology peer-to-peer group at Waterloo, and has been mentoring young girls and women of all ages for the past 3 years. She is also Adjunct Assistant Professor in the Pattern Analysis and Data Intelligence lab at the University of Waterloo. Dr. Hilal currently holds the position of Innovation Lead at Intelligent Mechatronic Systems.

### WHAT INCITED YOU TO JOIN IMS? WHAT IS YOUR ROLE AS INNOVATION LEAD?

Telematics is at the heart of IOT, and joining a company that has been in the business since 1999 and is currently leader in the telematics and IoT field was a great opportunity. Data intelligence and Innovation are at the heart of IMS's work right across the board.

I have been working at IMS for almost 4 years and currently hold the post of Innovation Lead. It has been a very exciting and interesting journey so far. The Innovation group's main role is to drive forward our products and technologies with tangible innovations that maintain IMS's leadership in these fields and guarantees we are providing our client base with the best possible services. We use machine learning to build engines for data mining and data intelligence, work on numerous proof-of-concepts, experiment with future technologies, and also put together pilot projects.

As Innovation Lead, I manage research collaborations with industry, academia and government partners. In addition to that, I lead internal projects for IoT and Big Data Intelligence. I report to the Vice President of Innovation who sits on the IMS board.

### WHAT IS IMS'S CORE BUSINESS?

We are a connected car company - that is our core business. We want to make cars smarter, greener and safer. A car doesn't have to be expensive in order to be safe. You can and should **get a safe, smart and environmentally-friendly car**, regardless of what make of car you are driving or what year it was produced. This has been our mandate from the very beginning, and what we have been working towards. We have a full suite of traveller intelligence solutions powered by IMS's DriveSync platform. These currently include:

**USAGE-BASED INSURANCE:** An insurance telematics solution enabling auto insurers to create uniquely targeted policyholder programs based on their needs and objectives, across both personal and commercial lines.

**PERSONAL TELEMATICS SERVICES:** A consumer-focused connected car solution delivering convenience, efficiency, productivity and safety services including young driver coaching, roadside assistance, remote diagnostics, emission monitoring and customized alerts.

**ROAD USAGE CHARGING:** A unique, in-vehicle system that uses telematics to eliminate the need

for expensive gantries and new physical road toll infrastructures. This system can help governments manage congestion as well as cover the cost of the road infrastructure and provide a fairer way of maintaining the road network than the petrol tax system, since it adopts the **Pay as you Drive model**: the more you use the roads, the more you pay for them. We are actually one of the biggest providers of **Pay as you Drive Road Charging** technology in North America.

**FLEET INTELLIGENCE:** A commercial fleet telematics solution offering vehicle tracking, vehicle health information and driver behaviour assessment. This fleet intelligence service is suited to both traditional fleets and synthetic fleets (rental agencies, car-share programs, dealerships, etc.), offering a broad range of relevant applications.

**DEALER INTELLIGENCE:** Empowering dealers with intuitive online and mobile interfaces to access vehicle assets, IoT and maintenance, and enabling them to secure recurring revenue while deriving CRM insights from connected vehicle data.

**IMS IOT MARKETPLACE:** The IoT Marketplace provides a broader set of consent-based sharing and data licensing opportunities across multiple telematics verticals and programs. This can help share program costs by allowing insurers to benefit from revenue sharing opportunities and offer a broader set of beneficial services to the policyholder.

**YOUNG DRIVERS INTELLIGENCE (YDI):** The Young Drivers Intelligence (YDI) service is tailored specifically for the new driver market, encouraging on-going dialogue between parents or supervisors and recently-qualified drivers, thereby helping improve driving behaviour and reinforce safe driving habits. Through this service, parents and supervisors get objective feedback and insight into the recently-qualified driver's behaviour, including performance measures based on rapid acceleration, speeding, harsh braking, sharp turns, excessive idling, driving during specific times of day, specific geographic regions, and anomalies in distance travelled.

### WHAT IS DRIVING THE CONNECTED CAR INDUSTRY? WHAT ARE THE CHALLENGES?

The term 'connected car' refers to its ability to connect with the outside world. The connected lifestyle we are living, accompanied by fast technological change, is pushing forward the need for connectivity in all aspects of our lives. Here



at IMS, we view the term 'connected car' a little differently, considering it to mean being connected to the driver, understanding what the driver needs and providing him with actionable information about those needs at the time.

With vehicles now coming equipped with the equivalent of 20 personal computers and processing about 25 GB per hour, **one of the key challenges is managing the humongous volumes of data that are being generated by all these vehicles and analyzing that data to extract useful insights.** The data is definitely characterized by its huge volume, variety, veracity and velocity, with large numbers of data points from different sensors and mobile devices offering varying degrees of accuracy being sent with high rates. As a general rule, vehicle data





is characterized by its varying precision, accuracy and data reliability. Another key challenge is mining insights from all this data, from understanding the vehicle dynamics to understanding the driver's intent and state of mind. IMS's extensive experience analysing driving and driver data provides significant value to the driver, parent, fleet manager, government, insurance carrier, and all partners across the transportation ecosystem. IMS's pattern analysis expertise also allows for a range of additional analytics solutions, from identifying driving events of interest, aggressive driving behaviour and even passive driver identification.

This is driven still further by the needs of Gen Z, who are more globalized and convenience-conscious than other generations. Gen Z do not only use one mode of transportation, they use bikes, shared cars, trains, the underground etc. As a result, the concept of mobility as a service has increased in importance. By viewing Mobility as a Service, **it is important to be mindful of the shift from vehicle ownership to vehicle usership.** We have been working to include this in our platform in an effort to provide an immersive experience regardless of what type of transport you are using, making the switch from the connected car concept to that of the connected driver, the connected lifestyle.

*AT IMS, YOU ARE DOING PIONEERING WORK IN THE CONNECTED CAR SYSTEMS FIELD. CAN YOU TELL US MORE ABOUT YOUR TECHNOLOGICAL ADVANTAGES?*

Our biggest advantage is that we have a full logistic solution that ingests data from different

sources, cleans it and applies data mining and intelligence algorithms to pull out all the useful insights that enable different verticals. This allows DriveSync to collect data from OBD II devices, or directly from vehicles which have connectivity, or even from smartphones and wearables. The OBD II dongle is a device that can be easily plugged into your car, as it simply slips into the diagnostic port, which is available in all cars. The OBD-II specification has been compulsory for all cars since the mid '90s. Another key advantage for our platform is the wide range of comprehensive verticals and value-added services that enhance the customer's safety and offer a more convenient driving experience.

We see **connected cars** as being cars connected to the driver, that understand what the **driver needs** and provide him with **actionable information** about those needs **at the time**

#### *HOW DO YOU CONNECT ALL THAT DATA TO BUSINESSES?*

The DriveSync platform has an extensive set of well-defined business-to-business interfaces to support the data and integration needs of both existing and future third party partnerships. The DriveSync platform includes interfaces based on open standards to simplify the integration of data from both third-party aftermarket and automotive OEM embedded telematics modules. **We are definitely an integration friendly company.** We also publish and deliver aggregate data insights to our partners, customers and potential customers through specialized reports, white papers and academic publications.

#### *INDUSTRY 4.0 AND PREDICTIVE MAINTENANCE IS A HOT TOPIC IN EUROPE. WHERE DO YOU STAND IN THIS FIELD?*

Industry 4.0 is actually what drives the IoT forward. People think connected cars are cars connected to the cloud. A car with WiFi would be a connected car, but we see **connected cars as being cars that are connected to the driver, that understand what the driver needs and provide him with actionable information about those needs at the time.**

For instance, if a car is low on petrol, we will provide details of the nearest petrol station and relevant insights like the current price of petrol and the time to destination.

Also, predictive maintenance is one of the value-added services we provide. By taking into consideration the vehicle's health, part wear and tear, driver behaviour, vehicle use and environmental conditions, we are able to predict the need for vehicle maintenance, the impact of such maintenance on the vehicle, and its estimated cost.

### *YOUR CAR TECHNOLOGIES AND SERVICES COMBINE TELEMATICS, INFOTAINMENT AND DATA INTELLIGENCE, BUT CAN WE ALREADY SPEAK ABOUT ARTIFICIAL INTELLIGENCE?*

IMS uses pattern analysis and machine learning to provide a range of analytic solutions from identifying driving events of interest, aggressive driving behaviour and passive driver identification. The DriveSync platform builds on years of innovation and advancements made possible thanks to IMS's collaborative research and development activities and established partnerships. **We are a Big Data company with terabytes of data points flowing every day, and we are working towards providing an even more comprehensive understanding of drivers and driving behaviour.** Our understanding of how people use their cars and how we can give them non-invasive feedback to make their drives safer, smarter and greener is definitely based on artificial intelligence and machine learning.

### *WHAT'S NEXT ON IMS'S ROADMAP? WHAT ISSUES WILL AFFECT CONNECTED CARS IN THE FUTURE?*

IMS is highly committed to continuous innovation. Our roadmap consequently has a large number of new features that are being released on a quarterly basis. Some of the key projects that are being released in the near future include:

**FIRST NOTICE OF LOSS TECHNOLOGIES:** Crash or incident detection and first-notice-of-loss (FNOL) applications for IMS solutions can rapidly provide customers with details of emergency contacts, towing services and other roadside assistance services following an incident, and can even report the incident to the authorities and insurers, if the driver so chooses, triggering immediate assistance. This technology leverages available device sensors (e.g. accelerometer, gyroscope), collected driving data (e.g. speed), and contextual data sources (e.g. road segment data) to detect crashes with high confidence.

**DISTRACTED DRIVING INTELLIGENCE MODULE:** IMS is taking an innovative approach to reducing distracted driving by providing a solution that seeks to understand how using a phone whilst driving actually impacts a driver's performance, and then shares this information with the driver via a personalized, fact-based appeal to which he can relate. IMS's Distracted Driving Intelligence quantifies the impact the user's distracted driving activity has on his driving performance and safety.

### *WHY IS RESPONSIBLE INNOVATION SUCH AN IMPORTANT PART OF YOUR COMPANY MINDSET?*

Responsible Innovation was first introduced during the Dutch Research Consul in 2006, and has been governed by an official EU framework since the EU's Italian Presidency in 2014. Humans are driven by ethics and morals, and the challenge is to embed both into the technologies we are building, especially when talking about autonomous cars. I believe that IoT companies should be conscious of the need to embed ethics, security, privacy, environmental-friendliness and democracy (access to technology with no discrimination) in their systems. Future generations will inherit the burden of whatever we build today. The technology we are designing today will have an impact on their world tomorrow. The key approach when it comes to responsible innovation is that of many hands, ensuring lots of different people work on each technology, with each one responsible for one specific part. But that creates an ownership challenge.

IoT companies should be conscious of the need to embed **ethics, security, privacy, environmental friendliness and democracy** in their systems

As a connected car technology company, IMS is aware of how important responsible innovation is, and it is something we have been working towards. IMS is ISO 9001 and 14000 registered, and ISO 27000 and ITIL compliant. Our data is all secured and private, we are trusted by governments and large insurance companies for our secure data hosting and security.

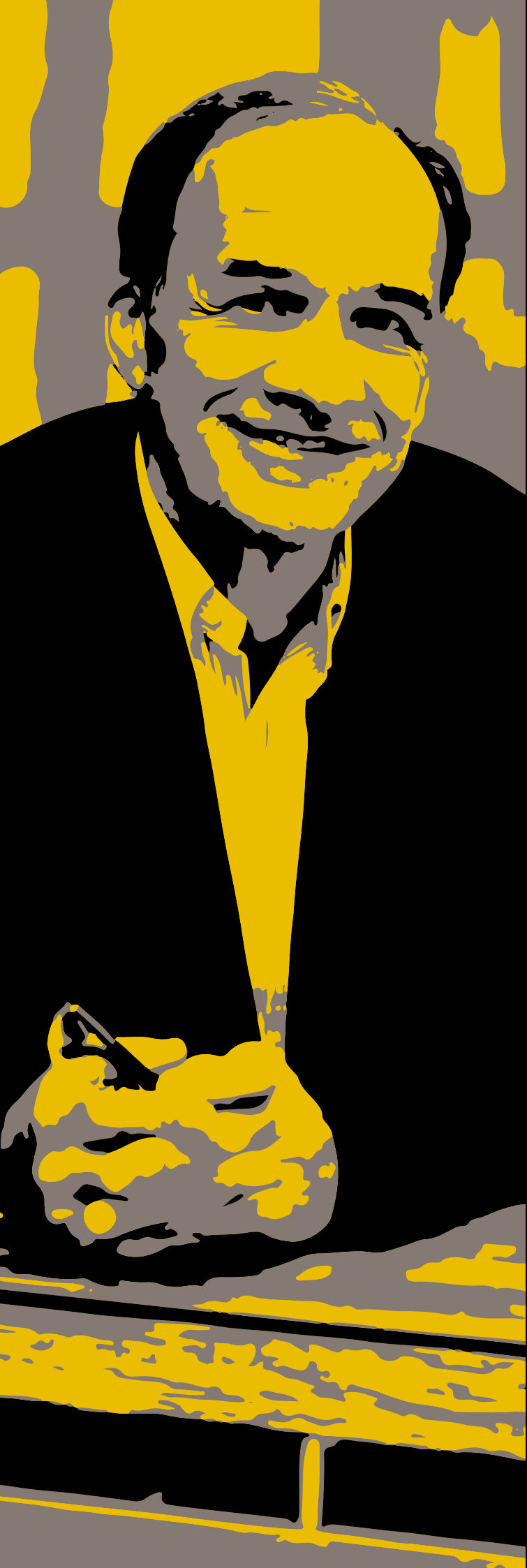
IMS's VP of Quality oversees the quality of operations in terms of the technology developed as well as the processes used. Our leadership team engages the IMS team to enhance the sense of product ownership. **Our Legal team also has very strict regulations in terms of data persistence and who owns the data.** The DriveSync platform builds on years of innovation and advancements made possible thanks to IMS's collaborative research and development activities and established partnerships.

### *YOU ARE INVOLVED WITH WOMEN IN TECH... HOW IMPORTANT IS IT TO GET YOUNG WOMEN INVOLVED IN THESE PROGRAMS?*

I believe that there is a global need for more women in technology and in general more diversity in the technology sectors. Having such diversity helps foster holistic product design. One good example is when vehicle manufacturers started developing the first airbag systems. These manufacturers had their team of engineers design the airbag system to provide additional safety features. However, when these airbag systems were deployed in the event of an accident, people with smaller body sizes, like women and children, got injured. They found that the design of the airbag system was based on the average size and weight of men, and they had forgotten to cater for the largest portion of vehicle users - women and children. Women may think, act, or approach problems differently. They tend to think not only of **women, but of men, children and the elderly.** Different perspectives can help build holistic solutions to the problems engineers are solving.

I believe that there is a **global need for more women in technology**

It is definitely important to build teams that are highly diverse. We need to design for a community, not for a single range of people, and encouraging diversity most definitely helps with that.



JEAN-YVES

**ROBIN**

DG

**OPENHEALTH COMPANY**



# Au coeur de la **collecte** et de l'**analyse** des **données** de **santé**

**Jean-Yves Robin**, médecin, il consacre sa carrière à la communication et aux technologies de l'information appliquées à la santé publique.

Il fonde en 1998 Uni-Médecine, qui deviendra Santeos, première entreprise française spécialisée dans les applications Internet au service de la santé. En 2008, le gouvernement lui confie la création d'une agence chargée de la politique publique en matière de numérique en santé, l'ASIP Santé, qu'il a dirigée pendant 5 ans. Il est également l'auteur en 2014 de « Santé : L'urgence numérique ».

Convaincu que l'analyse des données de santé est un puissant levier pour la transformation des systèmes de santé, il lance en 2015, avec Patrick Guérin, OpenHealth Company puis l'Institut OpenHealth de soutien à la recherche en Data Sciences de la santé avec le Professeur Marius FIESCHI.

Il est administrateur de la Fédération Nationale des Tiers de Confiance.

## QUELLE EST L'ORIGINE D'OPENHEALTH COMPANY ?

Autant l'internet était le sujet d'innovation de la fin du siècle dernier, autant les données constituent un des enjeux majeurs de celui-ci et particulièrement dans le secteur de la santé. Après cinq années passionnantes à la direction de l'ASIP Santé auprès du ministère de la santé, les conditions n'étaient manifestement plus réunies pour poursuivre les réformes en matière de santé numérique que j'avais initiées avec l'équipe de Roselyne BACHELOT. Je suis ainsi, par goût, retourné vers une aventure entrepreneuriale et ai logiquement choisi le sujet des données de santé.

**Fruits de la numérisation en cours, les données et leur exploitation vont être à la source des grandes transformations de demain. La régulation de ces transformations est du ressort des pouvoirs publics.**

En revanche, il est impératif que se développe en France un écosystème créateur de valeur, à l'origine de nouvelles connaissances et de nouveaux services.

La France est en retard dans ce secteur faute d'une ouverture volontariste des données et d'une politique suffisamment ambitieuse notamment dans les collaborations entre les secteurs publics et privés. Il faut les encourager voire les provoquer. Peu d'entreprises ont développé des activités dans ce secteur. Parmi elles, **la société Celtipharm** créée il y a une quinzaine d'années par Patrick GUERIN disposait d'un savoir-faire dans les traitements des données en temps réel et d'une bonne connaissance des données relatives aux médicaments. Une entreprise existante en recherche de croissance, un contexte et une vision partagée, il n'en fallait pas plus pour créer un nouvel acteur économique ; le projet OpenHealth était né.

La constance de mon action publique ou privée depuis plus de 15 ans consiste à mettre le numérique au service de la santé publique.

## QUELLE EST LA VOCATION D'OPENHEALTH ?

La vocation d'OpenHealth est simple : créer des informations utiles et de nouvelles connaissances à partir des données de santé. Cette vocation mobilise des données de toutes sources : des données de santé bien sûr mais aussi environnementales, socio-économiques etc, sur le territoire français ou autres. Nous traitons ces données selon des méthodologies variées, rétrospectives, en temps réel ou de façon prospective par le suivi longitudinal de larges cohortes populationnelles. Ces données sont évidemment anonymes et traitées dans le respect de la vie privée.

Notre modèle économique repose sur l'analyse et **nous plaçons pour une ouverture large des données qui, particulièrement dans le secteur de la santé, devraient constituer un bien commun dès lors qu'elles sont débarrassées de tout caractère nominatif bien entendu.** Nous réalisons ainsi des analyses à des fins de suivi de marché pour les industriels de santé mais également à des fins d'études épidémiologiques et de santé publique pour les autorités sanitaires et l'ensemble des acteurs de santé. Notre équipe, composée de professionnels de santé, biostatisticiens, épidémiologistes, collecte et traite chaque jour plusieurs millions de lignes de données.

Un des cas d'application des données que nous collectons concerne l'épidémiologie. Le suivi de la consommation de paniers de médicaments



mesurée en temps réel permet de suivre et détecter des phénomènes sanitaires multiples. Nous utilisons ces données pour suivre des pathologies chroniques ainsi que des phénomènes saisonniers tels que les allergies, les épidémies de grippe et travaillons sur des modèles prédictifs de ces phénomènes.

A partir des données de consommation de produits de santé, nous pouvons également faire d'autres types d'analyses, par exemple, sur les modalités d'usage de ces produits après leur mise sur le marché.

Elles peuvent être différentes de l'usage initialement prévu lors des essais pharmaceutiques. Le Médiateur en est un exemple. Suite aux études sur les dérives de son usage, la communication avec les prescripteurs est devenue plus rigoureuse sur les modalités d'utilisation du médicament.

Nous sommes en capacité de suivre et de mettre à disposition de nos clients industriels et institutionnels des données dites « de vie réelle », qui permettent de comprendre l'usage des médicaments en conditions réelles. Cela apporte des informations précieuses sur le comportement thérapeutique des patients : instauration, changement de traitement, bon usage, persistance au traitement, détection d'effets indésirables, comorbidités ....

Nous menons enfin des travaux de recherche. A ce titre nous avons fondé l'**OpenHealth Institute** en 2015, association à but non lucratif dotée d'un conseil scientifique indépendant et dédié au soutien à la recherche dans le domaine de l'analyse des données de santé. L'Institut OpenHealth mène des actions pour promouvoir les Data Sciences en santé à travers des bourses de recherche, des formations, éventuellement des mises à disposition de données à des chercheurs pour faciliter leurs travaux. L'Institut a lancé en 2016 un premier appel à candidature

pour des bourses de recherche. 40 projets ont été soumis par des chercheurs issus de grands laboratoires de santé publique et/ou de Data Sciences, parmi lesquels nous avons retenu 9 projets autour d'approches innovantes en matière d'analyse de données dans la santé publique. Une nouvelle campagne de bourses est en cours pour l'année universitaire 2016-2017.

## QU'EST-CE QU'APPORTE RÉELLEMENT LE BIG DATA AUJOURD'HUI DANS LE DOMAINE DE LA SANTÉ ?

Nous traitons des volumes conséquents de données (plusieurs millions de lignes par jour), mais le terme de Big Data est souvent galvaudé et ces volumes n'ont rien à voir avec certaines applications qui portent sur des millions de données à la minute, comme certains capteurs par exemple.

Le premier effet du Big Data c'est d'attirer l'attention sur les valeurs de la donnée. **Le Big Data met le projecteur sur des techniques de Data Mining certes anciennes mais totalement sous-utilisées dans la santé.** Ce qui est véritablement nouveau aujourd'hui, c'est certes la capacité à traiter de gros volumes de données mais aussi et surtout **le fait de pouvoir désormais réutiliser des données produites initialement à d'autres fins** (données produites à l'occasion d'actes de soins, données médico-administratives, données produites par les patients eux-mêmes...). Cette notion de « finalité compatible » développée dans le nouveau Règlement européen sur la protection des données personnelles ouvre des perspectives formidables. Avant, lorsque l'on réalisait une étude, on collectait des données de façon ad'hoc, sur la base de critères à partir desquels on remplissait les bases de données. Désormais les données produites pour des finali-

tés telles que soigner, vendre des médicaments, etc. sont réutilisées pour des études de santé publique par exemple. Il s'agit d'un nouveau phénomène, **le patrimoine de données existant peut être réutilisé pour des finalités d'intérêt général et compatibles avec la finalité initiale.**

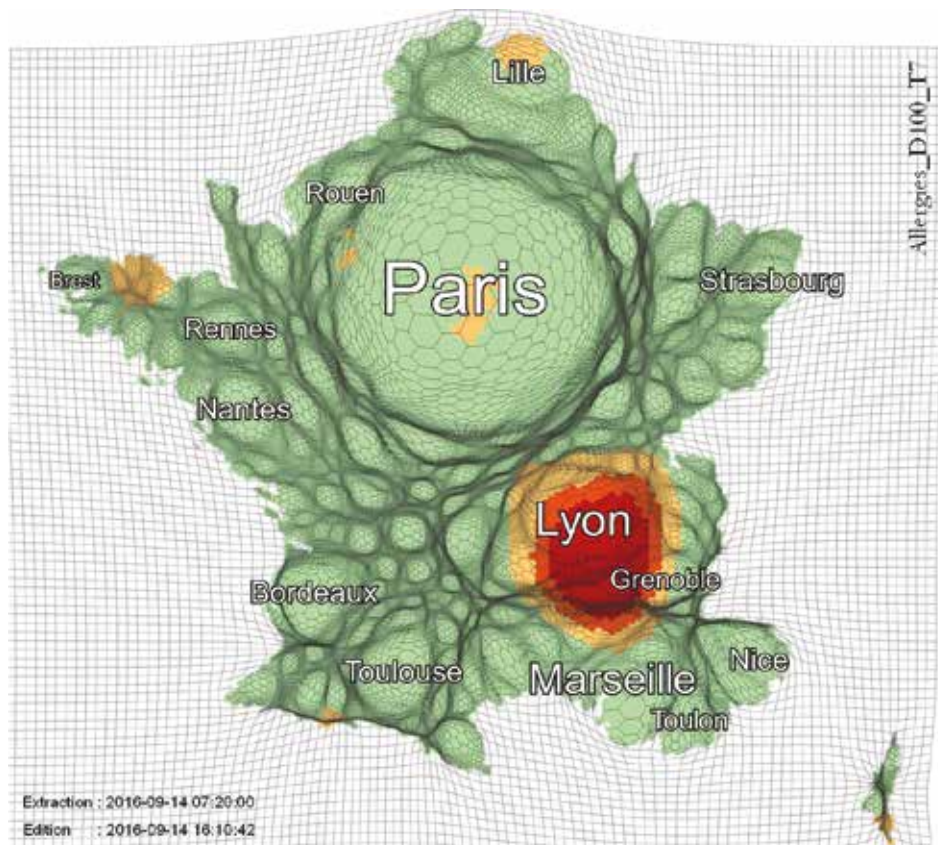
### LE PRÉDICTIF EST-IL D'ACTUALITÉ DANS LE DOMAINE DE LA SANTÉ ?

L'analyse des données est le support à la création d'algorithmes et de modèles prédictifs, afin de mieux évaluer un risque ou une probabilité de survenue d'un événement et d'en identifier les causalités en vue notamment de le prévenir. Pour exemple, une étude publiée au Canada a montré que l'utilisation des données permettait de prévenir la survenue d'infections néonatales chez les nouveau-nés, avant même l'apparition des premiers signes cliniques.

De même, l'analyse des données permet de prédire et donc possiblement de prévenir les réadmissions hospitalières pour des sujets âgés en particulier ; ces indicateurs se déploient aujourd'hui dans un certain nombre d'hôpitaux notamment nord-américains avec des incitations financières fortes.

De nombreux travaux sont menés sur les modèles de propagation des épidémies. La grippe est ainsi un des modèles intéressants qui peut être appliqué à d'autres phénomènes sanitaires. Chez OpenHealth, nous travaillons ainsi sur le développement d'algorithmes prédictifs en collaboration avec des laboratoires de recherche renommés, comme le laboratoire de Mathématiques appliquées de CentraleSupélec avec lequel nous avons conclu un partenariat.

### VOUS UTILISEZ DES CARTOGRAMMES POUR REPRÉSENTER LES PHÉNOMÈNES ÉPIDÉMIQUES, COMMENT CELA FONCTIONNE-T-IL ?



La Data Visualisation est un nouvel enjeu. Il s'agit d'une spécialisation à part entière, une véritable discipline qui est en train de naître et qui revêt un intérêt tout particulier dans la santé. Nous utilisons pour certains phénomènes une représentation cartographique par bassin de population. Le fond de carte classique est déformé proportionnellement à la densité démographique. Ainsi, si l'Ile-de-France compte 20% de la population française, elle occupera 20% de la superficie de la carte.

Cette approche utilisée en épidémiologie permet de mieux visualiser la propagation des phénomènes épidémiques ou sanitaires corrélés à la population et non à la surface du territoire.

Cela permet également de suivre en temps réel les personnes en fonction de leurs déplacements et d'introduire la notion de démographie dynamique. Il peut s'agir de phénomènes de fond comme les déplacements durables de population (des campagnes vers les villes par exemple) mais aussi de déplacements temporaires et courts comme les périodes estivales qui ont aussi leurs impacts sanitaires. C'est particulièrement intéressant en ce qui concerne les migrations de population. Dans le cas du virus Ebola, la première étape est de tracer les flux des personnes et les premiers cas d'épidémie pour comprendre l'origine de l'épidémie et comment elle se propage.

Au niveau du territoire, nous avons par exemple utilisé des données de téléphonie mobile pour étudier les mouvements de populations saisonniers. Un autre cas d'étude est celui des allergies. Les informations exploitées sont les données de délivrance de médicaments utilisés pour le traitement symptomatique des maladies allergiques. Une liste de médicaments est définie avec des médecins allergologues. Elle comprend principalement des médicaments antihistaminiques. Les données sont transmises anonymement et quotidiennement par le réseau des pharmacies paneuropéennes d'OpenHealth Company.

Les résultats sont fournis sous la forme d'un Indice. Un IAS Allergies de 100 signifie qu'il y a une absence d'allergie saisonnière. Un IAS

Allergies de 140 signifie qu'il y a 40% de plus d'allergie que le niveau de base. Les traitements statistiques sont effectués quotidiennement. La dynamique spatiale des manifestations allergiques est visualisée grâce à des cartogrammes iso démographiques.

Les données de vie réelle présentent un intérêt fort pour le pilotage des politiques de santé. Pour exemple, la couverture vaccinale des populations à risque contre la grippe est très en deçà du taux recommandé par l'OMS. Le suivi des données temps réel permet de mener des actions de communication sur les populations à risque et d'évaluer l'efficacité de chacune des actions. Il s'agit du b.a.-ba en marketing mais ce n'est pas encore appliqué concernant les politiques de santé. La culture de l'évaluation n'est pas encore suffisamment présente dans l'action publique mais la Data Analytics va changer le regard des citoyens et des décideurs.

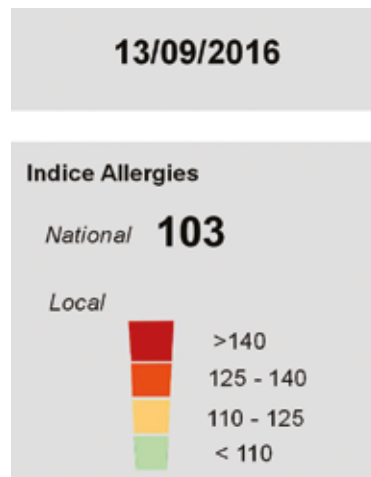
### QUE PENSEZ-VOUS DE L'INITIATIVE DU GOUVERNEMENT SUR L'OUVERTURE DES DONNÉES DE SANTÉ ?

Il y a un mouvement d'ouverture en Europe, le nouveau Règlement européen va dans ce sens avec notamment les notions de finalité compatible, de portabilité des données et la simplification des formalités préalables. L'Europe s'oriente véritablement vers une facilitation de l'accès aux données.

En France, on constate encore deux mouvements. La loi sur le numérique montre la volonté d'ouverture des données et d'amélioration de l'accès à ces données créatrices de valeur. La loi de santé de janvier 2016, qui souhaitait s'inscrire dans l'esprit de cette ouverture, notamment au travers de son article 193, n'est cependant pas à la hauteur des ambitions affichées. Elle clarifie certes certains points mais en pratique complexifie l'accès aux données. **Je pense que notre cadre juridique devrait être plus facilitateur pour nous permettre de faire face à la concurrence des autres pays européens et surtout celle des Etats Unis.** L'Etat en France se positionne trop souvent en acteur là où il ne devrait être que régulateur et facilitateur.

On parle souvent des mesures économiques pour soutenir la croissance. Mais **l'ouverture des données et la création d'un environnement juridique propice ne coûteraient rien et contribueraient significativement à la croissance par l'innovation.**

Représentation de l'allergie saisonnière à l'ambrosie dans le bassin de Lyon



## BENJAMIN CARLIER

Responsable d'incubateur  
de startups

**LE TREMPLIN**



# Les startups sportives à l'assaut du Big Data

**Benjamin Carlier** est diplômé d'un Master en management des organisations sportives. Chef de projet et chargé d'étude pendant 3 ans chez Carat Sport, agence de conseil en communication et marketing sportif, il a très vite choisi de se réorienter et est devenu le collaborateur parlementaire de Valérie Fourneyron à l'Assemblée Nationale avant de devenir Chef adjoint de son cabinet et conseiller sur un certain nombre de questions telles que le dopage et les relations avec les mouvements sportifs et le comité olympique. Suite au changement de gouvernement, il a décidé de rejoindre Paris&Co en Juillet 2014, l'agence de développement économique et d'innovation de Paris qui avait pour ambition, en lien avec la mairie de Paris, de créer un incubateur de startups dédié au sport. Il a été missionné dès 2014 pour une mission de pré-configuration de la plateforme d'innovation afin d'assurer qu'il y avait des startups intéressées pour rejoindre le projet mais également des partenaires potentiels de ces initiatives. Il est actuellement directeur du Tremplin. Nous avons souhaité avoir son point de vue sur les startups qui s'illustrent dans le domaine du Sport et de la Data.



## QUELLE EST LA VOCATION DU TREMPLIN ET QUEL EST SON RAYONNEMENT ?

Le Tremplin, lancé officiellement le 25 novembre 2014 a accompagné 17 startups pour la 1ère promotion, sélectionnées sur plus de 100 candidatures et 19 startups sur 120 dossiers lors de la 2ème édition. C'est un gage d'innovation et de réussite pour les startups sélectionnées. Le Tremplin, compte parmi ses partenaires membres fondateurs l'INSEP, la FDJ, Nike, l'UCPA, Unibail Rodamco, la Maif, les magasins Lepape, Accor Hotels Arena... Les partenaires font partie de la gouvernance du Tremplin et sont parties prenantes des décisions sur l'évolution du Tremplin. Etre partenaire du Tremplin leur permet également de faire du sourcing de startups et leur donne accès aux ateliers et conférences organisés à l'incubateur, ce qui leur permet de développer une culture de l'innovation en interne. **Le Tremplin crée un écosystème permettant aux acteurs du sport et de l'innovation de se rencontrer, d'échanger et de répondre ensemble aux problématiques liées à ce secteur dynamique.**

Le rayonnement du Tremplin dépasse les frontières de l'hexagone. **Il s'agissait d'une nouveauté mondiale que de créer un incubateur dédié au sport.** Ils ont reçu la visite de délégations brésiliennes, néerlandaises, italiennes, néo-zélandaises, espagnoles et canadiennes. Ce rayonnement est très intéressant pour les startups, car il leur permet d'être en contact avec des fonds d'investissements étrangers.

Le Tremplin propose un accompagnement individuel et collectif aux startups incubées et donne accès à un écosystème en leur proposant de participer à des événements, de pitcher devant des clients potentiels.

Dans le domaine de la data, le Tremplin propose des ateliers sur la thématique, des sessions de coworking, organisées entre les startups qui échangent et s'entraident sur les difficultés rencontrées.

Nous avons également mis en place la conférence annuelle DATA Foot avec la Fédération Française de Football (FFF) et HEC Alumni qui se tient au siège de la FFF. Cette conférence a abordé dans un premier temps la question de l'utilisation des données pour la performance puis, dans un second temps, la mise en valeur des données.

## QUEL EST LE NOUVEAU POTENTIEL DE LA DATA DANS LE SPORT ?

Il y a des sports dans lesquels les statistiques sont omniprésentes depuis des années comme le football américain, le basketball, le baseball etc.

**Les données ont toujours été omniprésentes dans le sport.** La mesure de la performance en athlétisme est basée par exemple sur une distance, une hauteur ou un temps... Dans les

Les données ont toujours été omniprésentes dans le sport

matchs de foot, on compte le nombre de corners, de tirs cadrés... Sur cette question, je conseille le film « Le Stratège », inspiré de faits réels qui raconte l'histoire d'une équipe de baseball de 4ème zone qui a gagné 20 matchs d'affilés, malgré des moyens limités, en recrutant son équipe uniquement en fonction d'analyses statistiques approfondies.



Exit le flair du recruteur, bonjour l'analyse statistique. Ce film a marqué un tournant dans l'usage des données dans le domaine du sport.

**On passe aujourd'hui des données loisirs, qui servaient des intuitions, à un véritable modèle de recrutement et d'analyse du sport.** Un virage a été pris sur la façon de faire du sport et notamment du sport de haut niveau. La multiplication des capteurs permet désormais de pousser l'analyse de la performance des sportifs de haut niveau à un niveau de précision exceptionnel.

Le Big Data est utilisé depuis quelques années pour optimiser la performance sportive, notamment avec l'exemple de l'équipe allemande accompagnée par SAP, lors de la Coupe du Monde de football 2014, qui a su se préparer parfaitement face au Brésil avec une victoire 7-1. Ce n'est pas le Big Data «qui va marquer des buts» mais «il permet d'analyser précisément les jeux et de prendre les meilleures décisions», expliquait Oliver Bierhoff, manager de la Mannschaft lors d'une interview réalisée en 2016 : «Les données seront de plus en plus importantes dans l'avenir. C'est pourquoi nous travaillons à l'académie pour qu'il y ait toujours plus d'experts en traitement des données dans le football.»

## LE BIG DATA A-T-IL UNE INFLUENCE DANS TOUS LES DOMAINES SPORTIFS ?

Dans le sport de haut niveau, de très nombreuses startups vont s'illustrer. La thématique des capteurs est extrêmement importante et **MAC LLOYD SPORT**, l'une d'entre elles, incubée au Tremplin, développe et fournit des capteurs pour équiper le monde du rugby. En match, les capteurs permettent d'extraire les informations sur les distances, les appuis au sol, la vitesse d'exécution des actions, les baisses de régime des joueurs, le rythme cardiaque, les séances d'entraînement, la charge de travail imposée au joueur ainsi que la prévention de

certaines blessures. On est par exemple capable de connaître le déséquilibre entre la jambe gauche et droite d'un joueur lors du match ce qui permet de détecter un signe précurseur de blessure. Aujourd'hui les joueurs de rugby vont regarder les données du match pour

On passe aujourd'hui des données loisirs, à un véritable modèle de recrutement et d'analyse du sport

mieux se connaître, mieux quantifier leurs réalisations et comparer leurs performances.

**ARIONE0**, dans le domaine de l'équitation, développe des solutions connectées d'analyse de la performance et de la santé des chevaux athlètes, un domaine dans lequel on manquait cruellement d'informations. La société s'appuie sur son expertise en physiologie équine et Data Mining pour assurer l'efficacité de ses algorithmes au service d'un tracking de l'activité physiologique et sportive des chevaux.

**CONNECTED CYCLE** s'illustre dans le cyclisme en concevant et développant des solutions de géolocalisation et de suivi d'activité à partir d'objets connectés. Connected Cycle a présenté la première pédale connectée au monde, alimentée par l'énergie du pédalage. La pédale alerte le propriétaire sur son smartphone lorsque son vélo est déplacé, et permet de le géolocaliser à tout instant. Elle enregistre automatiquement chaque trajet effectué : itinéraire, distance parcourue, vitesse, dénivelé, calories consommées. Les statistiques sont envoyées dans le cloud, et mises à disposition des utilisateurs à travers l'application. Elle s'accompagne aussi

d'outils destinés aux professionnels de la location pour mieux gérer leurs flottes de vélos.

**FOOTOVISION**, grâce à un logiciel de tracking à partir de vidéos de matchs de football, calcule toutes les données de position des joueurs, du ballon et les événements de jeu (passes, tirs, possession de balles etc...). Le logiciel les sauvegarde également en base de données. L'analyse peut être effectuée soit via une application représentant le pitch virtuel en 3D, soit directement sur la vidéo originelle en réalité augmentée. Ils sont capables de faire l'ensemble de ces analyses à partir du simple flux télé. Nous sommes abreuvés de chiffres à la télévision mais ceux-ci ne reflètent pas forcément la performance des joueurs. Ils ont développé un algorithme qui va chercher à détecter les joueurs, les coéquipiers autour de lui, la position des joueurs adverses, si les passes sont considérées comme faciles ou difficiles et à combien de joueurs ils peuvent potentiellement faire la passe. Footovision a pu mettre ses technologies et applications en tests dans plusieurs stades lors de l'Euro 2016.

**AMISCO & PROZONE** est un pionnier dans le suivi des athlètes, l'analyse des performances sportives et la gestion des actifs. Ils permettent notamment aux responsables et staffs techniques d'optimiser leur processus de décision avec des informations objectives et d'améliorer la gestion d'effectif et la stratégie de recrutement dans un contexte compétitif.

### ENTRAÎNEURS ET DATA SCIENTISTS : COMMENT SE POSITIONNENT-ILS ?

C'est un **business** dans lequel l'arrivée du Big Data permet de **diminuer l'aléa** et le **risque**

Le club du Leicester dispose d'une data room composée de quatre Data Scientists qui analysent en temps réel les données des joueurs. Il s'agit d'un métier d'avenir dans le football pour permettre de fournir des informations complémentaires à la mi-temps des matchs grâce au Big Data. Leur rôle va être de plus en plus important et les entraîneurs doivent apprendre à travailler avec eux pour des conseils performants. Néanmoins, il y a 2 ans lors du match de rugby France - Angleterre se déroulant au Stade de France, à la 65ème minute de jeu, la France perd de plus de 10 points. L'entraîneur anglais voit alors sur sa tablette que son numéro 9 est fatigué et décide de le remplacer. Il s'agissait d'une erreur de l'entraîneur car le demi de mêlée avait une influence irremplaçable sur le match et sur l'arbitre. La France va remporter le match. **C'est un business dans lequel l'arrivée du Big Data permet de diminuer l'aléa et le risque. Il s'agit d'une recette**

**miracle, en revanche si tous les clubs s'en emparent, le rôle de l'entraîneur en collaboration avec les Data Scientists sera toujours de gérer les hommes et de gagner des matchs par ses intuitions complémentaires aux datas collectées.**

### QUID DES DONNÉES DU SPORT AMATEUR ?

Le « Quantified self » prend de plus en plus d'importance, les données font également partie du quotidien des sportifs amateurs.

La startup **FOOTBAR** développe, par exemple, des capteurs connectés dédiés au football et fournit des statistiques détaillées des championnats amateurs.



**STRAVA** est historiquement leader dans l'analyse et la comparaison des performances des runners et cyclistes entre eux. Adopté par plusieurs millions d'utilisateurs, Strava collecte les données via des technologies GPS et fournit des analyses précises notamment sur les tronçons empruntés par les cyclistes permettant aux amateurs de se comparer aux professionnels. Le monde amateur alimente le monde professionnel et inversement.

La startup **MOJJO** s'illustre dans le tennis. Lors d'un match de tennis professionnel, on dispose de nombreuses données : nombre de points gagnés, zones de jeux. Ces données sont également extrêmement intéressantes pour des joueurs amateurs et permettent de décider quel devrait être le schéma idéal de jeu. Mojjo met ainsi à disposition une feuille de match détaillée et une bibliothèque vidéo de tous les événements du match. Leur technologie représente une vraie rupture dans la mesure où l'algorithme de détection n'utilise qu'une seule caméra, là où les concurrents en utilisent 5 ou plus. Dans chacun des sports, les services et analyses proposés sont sectorisés.

Dans le domaine amateur, on peut se demander si l'utilisation des données relève plus d'une mode ou d'un besoin de fond.

En revanche la Gamification du sport reste un des intérêts croissant lié à l'utilisation de la data. **RUNNING HEROES** collecte par exemple l'intégralité des données, à partir des objets connectés utilisés lors du parcours, pour récompenser les coureurs en fonction de leurs performances.

### UNE OPTIMISATION DES STRATÉGIES SPORTIVES EN TEMPS RÉEL EST-ELLE UTOPIQUE ?

L'optimisation des stratégies sportives en temps réel existe déjà. Dans le domaine de la Formule 1, les techniciens décident à partir des données qu'ils reçoivent en temps réel, de la nécessité de passer au stand ou non. La Data room de Leicester fournit des informations en quasi temps réel, les informations fournies sont dupliquées sur le même match, de même en rugby et en cyclisme.

En football américain, le Quarterback conseille directement à ses joueurs la stratégie en temps réel.

La question qui se pose est de savoir si le football en temps réel est possible. L'évolution de la réglementation va être déterminante. Pour exemple, les possibilités de donner des consignes au pilote de Formule 1 sont très limitées aujourd'hui. Les Fédérations nationales et les gouvernements vont être décideurs sur l'utilisation future de la data dans le sport.

La collecte des données spectateurs va également permettre de mettre en place des stratégies en temps réel, basées sur la connaissance. **DIGIFOOD** et **TECH4TEAM** s'illustrent dans le domaine.

**DIGIFOOD** propose aux spectateurs de sélectionner leur événement (matchs, concerts, universités, etc.) et de renseigner leur localisation pour passer commande et se faire livrer à leur place. Ils proposent ainsi un service supplémentaire pour l'expérience spectateurs mais récupèrent aussi des données utilisateurs qui vont pouvoir être utilisées par les marques en quête d'améliorer l'expérience et de personnaliser la relation client. L'analyse des données permet de mieux cibler, de mieux vendre et d'optimiser les revenus afin de rendre le spectacle plus attractif.

**À terme, on pourra envisager du marketing émotionnel.** Si les spectateurs sont connectés, on va pouvoir connaître leurs émotions, par exemple liées à un but de Zlatan, et proposer en temps réel aux spectateurs des objets marketing à la vente.

**L'un des derniers domaines d'avenir est celui du sport bien-être et santé,** les gouvernements le prennent en considération. L'entreprise **LSEE** étudie par exemple l'impact de l'activité physique sur notre métabolisme en permettant de déterminer l'alimentation et l'activité physique adaptées aux spécificités génétiques et métaboliques de chacun, grâce un suivi autonome et en temps réel de l'activité métabolique. **La personnalisation de la médecine sportive et l'individualisation de la pratique sportive pour perdre du poids est l'une des pistes d'avenir pour le Big Data dans le domaine du sport.**





## KÉVIN. VITOTZ

Fondateur et Président  
**TECH'4'TEAM**



# La Data fait son show dans la Billetterie Événementielle

Tous deux diplômés de l'ENSAE ParisTech en mathématiques appliquées et d'HEC, **Kévin Vitoz** et Ludovic Bordes ont souhaité mettre leurs compétences au service du marché de l'Entertainment en développant un outil permettant l'introduction du Big Data dans les enceintes sportives et culturelles.

C'est donc fraîchement diplômés, que ces deux passionnés de sport et férus d'analyse de données statistiques se sont lancés dans l'aventure entrepreneuriale en créant Tech'4'Team.

### PRÉMISSSES DE LA SOCIÉTÉ

La première année, nous avons travaillé en tant que consultants pour le Club de Saint Etienne. L'analyse des données des abonnés devait permettre de diminuer le churn (la perte d'abonnés) et d'augmenter la conquête de nouveaux clients. Nous avons également travaillé sur le scoring de ces abonnés et sur la récupération de l'ensemble des données des musées, boutiques et boutiques, que nous avons nettoyées, enrichies et structurées afin de permettre à l'organisation de s'adresser à ses clients de manière plus ciblée. Dès l'été 2014, nous avons commencé à travailler avec de grands clubs sportifs en leur apportant nos compétences sur des problématiques liées à la billetterie, avons embauché nos premiers employés et débuté le développement de notre logiciel et produit SAAS (Software As A Service) d'optimisation des revenus billetterie dans le secteur de l'Entertainment.

### COMMENT ALLIEZ-VOUS BIG DATA ET ENTERTAINEMENT ?

Nous avons d'abord cherché à **rendre accessibles à des acteurs du spectacle vivant, qui ne disposent ni de compétences datas, ni de temps pour travailler les données, des analyses adaptées à leur contexte et leurs métiers via de la Datavisualisation simple et agréable**. Nous collectons des données externes telles que la météo, la performance sportive, le calendrier scolaire, la notoriété des sportifs et des équipes, les grèves... Nous les croisons ensuite avec toutes les données historiques de vente afin de fournir des prévisions d'affluence ainsi que des recommandations intelligentes qui permettront aux structures d'optimiser leurs ventes. L'outil de reporting fourni à nos clients est couplé d'un CRM prédictif, qui permet d'enrichir le profil des clients grâce à des données exogènes. Par exemple, à partir du prénom, nous

serons capables de trouver le sexe et l'âge d'un client ou encore d'évaluer la rémunération à partir de son code postal.

La récupération des données lors du processus d'achat est très contraignante pour les organisateurs d'événements. **L'enrichissement prend alors tout son sens en réduisant le nombre d'informations demandées aux clients et en augmentant la connaissance que nous pouvons avoir d'eux**. Cet « effet ciseau » permet alors la mise en place d'opérations marketing ciblées et le développement de stratégies de fidélisation plus efficaces. Par ailleurs, nous sommes directement connectés aux outils d'emailing, de SMS et de Call Centers pour récupérer des informations sur les clients et les intégrer au dispositif.

Nous avons ensuite développé un outil incorporant davantage de techniques d'exploitation de la Big Data. Il s'agit d'un **instrument de pilotage des revenus billetterie et de tarification innovante qui permet par exemple de calculer le nombre de billets à allouer à chaque distributeur, le nombre de places à attribuer à chaque catégorie et par-dessus tout de proposer le bon prix au bon client, au bon moment, par le bon canal de distribution**. Ces calculs sont effectués grâce à des **techniques de Yield Management** et à des **prévisions en temps réel**. Le Yield Management regroupe de nombreux acteurs dans le secteur de l'aérien et du tourisme, mais leurs modèles mathématiques sont encore trop peu appliqués au marché de l'Entertainment.

### QUELLES VARIABLES SONT PRISES EN COMPTE ?

Nous sommes en mesure de récupérer et traiter les données billetterie et CRM des structures de l'Entertainment. Par des accords variés avec les différents acteurs billetterie comme **WEEZEVENT**, **SÉCUTIX** ou **APARTÉ**, notre outil agrège et reformate les données qui en sont issues.

Nous intégrons également les données issues des réseaux sociaux ce qui nous permet de faire le lien entre les acheteurs de billets et les communautés (fans, followers de certains artistes). Nous sommes ainsi capables d'identifier certains profils de clients, de permettre un démarchage en un clic et la diffusion d'une communication ciblée.

Nous collectons également des données externes comme les performances sportives à jour qui ont un très fort impact sur l'achat de billets. Pour les spectacles, nous récupérons les données concernant la notoriété des artistes sur des sites en open source. Nous récoltons grâce à des APIs le nombre de fans et de followers, les données en open source de sites comme **SONGKICK**. Enfin, nous travaillons avec des startups comme **GUEST VIEWS**, livre d'or numérique dans le domaine culturel, qui nous permet de collecter, d'analyser et de valoriser les données et avis des visiteurs ou clients. Toutes ces informations sont utilisées pour l'enrichissement des données afin d'améliorer la connaissance client.

### QUID DES ALGORITHMES BIG DATA ET DE VOTRE BUSINESS MODEL ?

Notre algorithme d'optimisation des revenus billetterie a été développé en interne. Nous sommes Data Scientists de formation et nous tenions à développer des modèles algorithmiques qui soient internalisés. **Plus nous avons de clients plus nos modèles apprenants sont performants.**

Nous débutons avec une version d'essai gratuite et évoluons vers un modèle Freemium, ce qui est assez innovant pour un marché aussi peu mature sur ces questions. **Nous nous inscrivons donc dans une stratégie d'éducation du marché pour en tirer les bénéfices.**

En France, nous faisons partie des pionniers du Dynamic Pricing dans l'Entertainment. Notre objectif est de démocratiser le Big Data dans ce secteur, dont certains acteurs comme les musées, clubs de sport ou parcs d'attraction sont déjà convaincus des bénéfices. Nous travaillons notamment avec **la Ligue de Football professionnel, le Musée Grévin, la Cité des sciences, le Théâtre du Chatelet, le Weather Festival et Les Papillons de Nuit** pour l'optimisation de leur billetterie.





# AMÉLIE DELOFFRE

Data Storyteller

SPORT HEROES GROUP



## Quand Data & Gamification entre dans la course

**Amélie Deloffre** a commencé sa carrière en tant que Data Analyst & Consultante en visualisation de l'information au sein de l'agence Publicis Consultants. Elle y a apporté dans un premier temps sa connaissance de la data visualisation, à laquelle elle a formé les collaborateurs en interne tout en améliorant, en parallèle, les différents produits proposés par l'agence (études, audits d'image, cartographies, reporting...). Désireuse d'aller plus loin dans la data, elle a fondé en interne DATACRAFT, une cellule d'experts dont le but est de mettre les données (internes, externes et social média) au coeur des dispositifs de communication des grandes marques. Après un an passé à évangéliser le marché auprès d'entreprises du CAC 40 encore peu matures sur le sujet, elle se lance dans l'aventure start-up et rejoint Sport Heroes Group (ex-Running Heroes) qui collecte les données de milliers de sportifs. Un nouveau défi alliant deux passions : le sport et la data.

Heroes Group (ex-Running Heroes) est né de l'envie des deux fondateurs, Boris Pourreau et Jean Charles Touzalin, d'encourager la pratique sportive en trouvant de nouveaux leviers de motivation tel que le système de récompense, la gamification ou le storytelling. L'envie de ramener le sport à ce qu'il était l'origine pour chacun d'entre nous : un jeu, loin de l'aspect purement performance qui est très présent aujourd'hui dans le sport connecté. D'où notre crédo actuel : «**Live sport differently**».

### QUEL EST LE CONCEPT, SERVICE PROPOSÉ ?

Notre startup a été créée il y a deux ans. Activité originelle de la startup, Running Heroes est une plateforme dédiée à l'encouragement des coureurs amateurs, en les faisant bénéficier de récompenses et de réductions chez quelques 350 marques partenaires telles que Nike, Spotify, Adidas, Uber, Bio c'Bon ou encore Sarenza. **Le principe : plus les sportifs courent, plus ils gagnent de point à échanger contre les récompenses de leur choix.** Plus récemment, nous avons lancé deux nouvelles activités (We Are

Heroes et United Heroes) basées sur la création d'expériences connectées : des courses dématérialisées permettant d'engager un grand nombre de participants, partout dans le monde et ce, grâce à la data émise par les applications et montres GPS.



### QUELLE UTILISATION DE LA DATA FAITES-VOUS AU SEIN DE SPORT HEROES ?

D'un point de vue purement data, notre plateforme Running Heroes permet aux annonceurs d'accéder à une audience ultra-ciblée. **Nous proposons des campagnes permettant d'affiner encore davan-**

**tage le ciblage et ainsi d'adresser le bon message à chaque type de runners en fonction de critères sociodémographiques ou comportementaux, liés à leur pratique.** Nous offrons, par exemple, à une marque comme Nike la possibilité de promouvoir une paire de chaussures de casual running sur les racers qui s'entraînent une fois par semaine.

Plus globalement, **nous envisageons le Big Data dans une visée de meilleure compréhension de la pratique sportive connectée.** Nous allons notamment créer un Observatoire du running qui rendra compte des usages des sportifs en France : qui sont-ils ? quand et comment courent-ils ? Cet enjeu de compréhension est important pour l'ensemble du monde sportif, mais également pour les pouvoirs publics qui souhaitent aujourd'hui prendre en compte la pratique sportive dans l'aménagement du territoire.

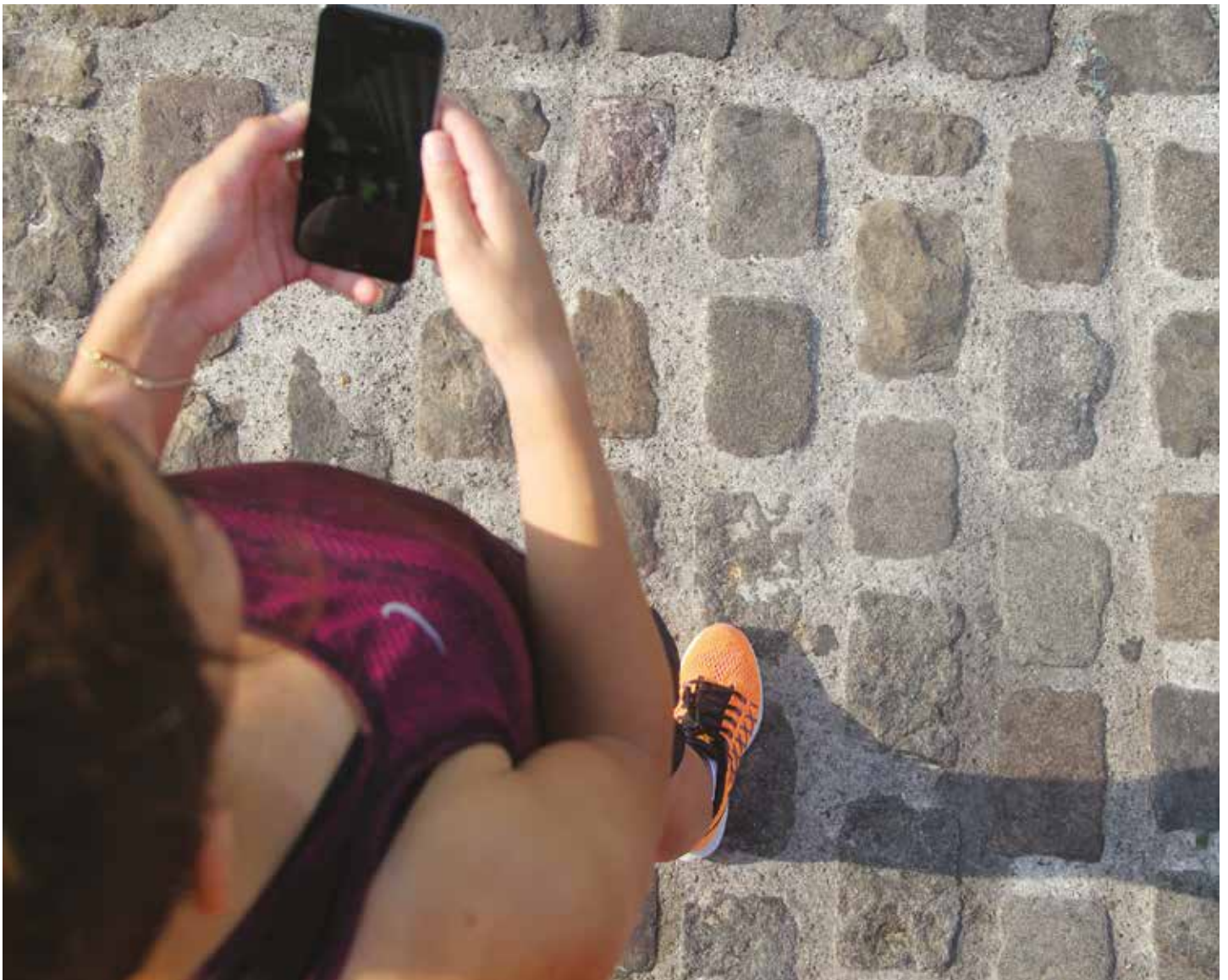


### QUELLES DONNÉES RÉCOLTÉES ?

Nous récoltons donc des **données d'activité sportive, des données sociodémographiques telles que le sexe, l'âge du sportif et la localisation ainsi que des données comportementales à savoir les distances parcourues, la vitesse de course, le dénivelé, le jour et heure de la course**. Nous utilisons des bases de données classiques gérées par nos développeurs en interne. Nous n'avons pas de « Data Scientists » à proprement parler au sein de l'équipe. En revanche nous nous appuyons sur un cabinet de conseil pour la gestion de SAS Visual Analytics afin de pluggier l'ensemble des données dans un Dashboard utilisable par l'ensemble de nos équipes.

### QUEL AVENIR POUR LA DATA CHEZ SPORT HEROES GROUP ?

**La data est le coeur du réacteur de Sport Heroes Group.** Toutes les précieuses données que nous stockons et analysons en ce moment même nous permettront d'acquérir une place centrale dans le monde du sport avec une connaissance très fine et unique des sportifs. **Les usages de nos clients nous poussent constamment à développer de nouveaux services alliant challenge, gamification, data et mêmes des aspects caritatifs.**





## ELIAS HADAYA

Director of Insight and Research

**VISION DU MONDE CANADA**



# Customer Intelligence et Big Data au **service** d'une **ONG**

**Elias Hadaya** a grandi en France et après les Ecoles Préparatoires, il a obtenu un diplôme de DEA en Econométrie et un Magistère d'Ingénieur Economiste. Il a travaillé pendant une quinzaine d'années dans le milieu bancaire et des télécommunications. Son dernier rôle, occupé pendant plus de 8 ans, était celui de Vice-Président adjoint Consumer Insight. Il avait la responsabilité des Databases marketing, du Data Mining et était également pilote d'affaires pour déployer la solution CRM à travers toutes les succursales bancaires.

Il a été approché pour un poste au sein de Vision Mondiale Canada dont le siège social est en Ontario à Toronto. Sa femme et lui soutenaient l'organisation depuis plusieurs années et il était au Conseil d'administration d'une autre ONG qui venait en aide aux jeunes défavorisés. Il a décidé de rejoindre Vision Mondiale Canada surtout en raison de la cause humanitaire, de leur plan d'affaires, et de la possibilité offerte de mettre le domaine du Big Data et de l'Analytics au service d'un organisme de bienfaisance qui sauve des vies.

## QUELQUES MOTS SUR VISION MONDIALE CANADA ? LE BUT DE L'ORGANISME ET SES CHALLENGES ?

La mission de l'organisation est de travailler à travers le monde dans les pays les plus en difficulté en termes de développement international. Nous travaillons dans 255 communautés à travers 48 pays en lien avec les leaders locaux, les communautés, les familles et les enfants. Vision Mondiale Canada fait une levée de fond d'à peu près 400 millions de dollars (CDN) par an et vient en aide à près de 500 000 enfants à travers le Monde. Vision du Monde internationale réalise une levée de fonds de près de 2,7 milliards de dollars à travers une centaine de pays. Il s'agit d'un organisme de bienfaisance d'envergure, une large organisation avec une grande capacité à utiliser les données.

**Nous sommes la voix de ceux qui n'ont pas de voix**

Quand nous commençons une mission d'aide, cela prend à peu près 10 à 15 ans pour rendre le village et la communauté indépendants. Nous évaluons les besoins et agissons sur de nombreux secteurs qui sont la fondation de la survie humaine tels que l'éducation, la santé, le développement durable, la nutrition, l'agriculture... Nous faisons également un travail de lobbying et d'Advocacy auprès des gouvernements, pour la lutte contre l'esclavagisme et les abus sur les enfants dans les pays en voie de développement pour aider au développement durable et éthique. Nous travaillons également auprès des gouvernements de pays développés comme le Canada ou la France, l'Union Européenne et les Nations Unies. **Nous sommes la voix de ceux qui n'ont pas de voix.** Nous travaillons enfin avec le World Food Program pour être certains de pouvoir amener le développement et la voix des médias, là où ils sont nécessaires.

### L'ÉQUIPE ANALYTICS DE VISION DU MONDE

Je dirige l'équipe recherche et Big Data canadienne composée de 18 personnes, qui gèrent le descriptif, la production de rapport et Dashboards... jusqu'au prédictif. « **We cover a wide range of projects, but we are also deep** »,

nous ne pouvons plus nous contenter d'un seul modèle prédictif, l'horizon de travail devient de plus en plus large et de plus en plus complexe. Une partie de l'équipe est spécialisée en statistiques, mathématiques et Data Science et l'autre partie est composée de personnes qui ont un profil très affaires. Recruter représente une véritable difficulté pour Vision du Monde car nous sommes régis par l'équivalent canadien de la Direction des Finances Française et nous ne pouvons utiliser plus d'un certain ratio pour la levée de fonds et l'administratif. Il est donc très difficile d'attirer les bons talents. **Un de mes conseils serait de pas tomber dans le piège de chercher uniquement des superstars pour former une telle équipe mais de se doter de bons spécialistes dans leur domaine d'affaires.**

**We cover a wide range of projects, but we are also deep, nous ne pouvons plus nous contenter d'un seul modèle prédictif, l'horizon de travail devient de plus en plus large et de plus en plus complexe**

### COMMENT AVEZ-VOUS CONSTRUIT VOTRE EXPERTISE EN DATA ANALYTICS AU SEIN DE VISION DU MONDE ?

Quand je suis rentré en poste, je me suis rendu compte qu'il s'agissait d'une équipe forte, mais avec un questionnement fort sur la valeur apportée par l'équipe. Tous les membres de l'équipe de direction ont déclaré qu'il s'agissait d'un véritable avantage compétitif d'avoir une équipe dédiée au Big Data, en revanche peu d'entre eux ont réussi à me décrire la valeur unique de l'équipe pour l'entreprise. J'ai donc commencé par positionner l'analytique, comme un véritable domaine d'affaire. C'est une pensée qui est très présente en France mais pas au niveau nord-américain. **Si c'est un domaine d'affaires, un métier, il faut le traiter en tant que tel, il faut investir dans les bonnes ressources, dans les données, dans la tech-**

**nologie et être maître et leader des décisions prises au sein de son équipe.** Nous sommes ainsi passés d'une utilisation très tactique de l'information à une utilisation très stratégique, nous avons fait de la segmentation, du predictive modeling, du digital analytics, afin d'étudier les réactions des clients sur nos outils internet etc.

Les données sont primordiales, tout d'abord pour comprendre les personnes avec qui nous travaillons pour faire la levée de fonds, comprendre le marché, les canadiens et ce qui motive leur décision de faire un don. L'analyse des données nous permet de mettre en place un targeting ciblé pour des actions efficaces sur le bon segment et pour la bonne initiative.

**Les données sont également indispensables pour améliorer l'expérience client et offrir en contrepartie des dons, la possibilité de faire partie d'un organisme qui sauve des vies, d'un groupe qui a de l'influence.** Dans le passé, nous n'avions pas de traitement personnalisé pour nos donateurs. Nous sommes passé d'un traitement marketing « taille unique » à un traitement marketing personnalisé basé sur les caractéristiques propres à chaque segment.

Nous vivons dans un Monde où les clients vont sur Amazon et profitent d'une très bonne expérience digitale et utilisateur. Nous sommes une oeuvre de bienfaisance, avec moins de possibilité qu'Amazon ou Ebay, néanmoins nos donateurs, quand ils visitent notre site web ou interagissent avec nous, s'attendent à la même expérience et la même qualité d'échange. Il est donc impératif de les comprendre, les rencontrer là où ils sont pour pouvoir faire des levées de fonds efficace.

Enfin, nous souhaitons utiliser les données pour faire une programmation de l'aide efficace. Il y a énormément de projets et les données, telles que le nombre d'enfants aidés, la localisation des communautés aidées, sont gérées par les responsables de programme directement, de même pour les KPIs et Dashboards qui présentent ces éléments. Notre équipe commence à s'y intéresser et à vouloir amener plus de sophistication dans ce domaine, mais nous en sommes encore aux balbutiements sur l'utilisation du Big Data dans ce domaine.

### COMMENT ÊTES-VOUS POSITIONNÉS PAR RAPPORT À VOTRE DIRECTION ?

Je reporte directement au Chief Marketing Officer. Elle reporte, elle-même, directement au président. Il n'y a donc qu'un niveau entre moi et le président et cela positionne la data et mon équipe à un niveau très stratégique. Nous avons collecté toutes les données et avons regardé les **Survival Analysis** de tous les donateurs pour appréhender quel était le risque d'annulation de leur aide. Nous avons ainsi construit un modèle de **Forecasting** des dons sur les cinq prochaines années qui a été présenté au conseil d'administration. Ce modèle permet de piloter les activités pour savoir dès qu'une action est lancée, quel en sera l'impact dans 5 ans et quelles seront la durabilité et rentabilité des projets.

J'ai également revu la structure de l'équipe en 2012. Avoir des gens qui connaissent les données, capables de les forer et de les analyser est primordial, mais il est tout aussi important d'avoir des personnes, qui ont un Business Acumen, une très bonne connaissance des affaires et de la levée de fond. J'ai







donc décidé de créer des Business Partners, ce sont des partenaires d'affaires internes pour tout le Marketing. Ils sont dans les rencontres d'équipes des autres divisions afin d'écouter les besoins, ce qui leur permet d'être vraiment connectés au business et d'avoir une approche plus stratégique, lorsqu'ils développent des projets. N'engager que des Data Scientists est un vrai danger. Ne pas avoir de gens d'affaires au sein d'une équipe analytique, c'est prendre le risque que le pouvoir réside dans les personnes qui ne comprennent pas l'analytique et le data. Nous sommes partis d'une logique, dans laquelle on nous demandait de fournir un chiffre, pour aboutir à un rôle de consultation de l'équipe sur des problèmes d'affaires.

N'engager que des  
**Data Scientists** est un  
vrai danger

### QUELLES TECHNOLOGIES UTILISÉES ? DU PRÉDICTIF ? DES MODÈLES DÉVELOPPÉS EN INTERNE ?

Nous faisons du descriptif, du reporting, nous utilisons notamment des outils IBM dernier cri mais nous avons aussi tout un pan prédictif : nous avons en effet développé des modèles pour savoir quelles personnes cibler avec quelle offre mais également pour prédire qui va arrêter son soutien aux enfants. Pour ce dernier cas, **nous nous appuyons sur un modèle de survie ou Survival Model.**

Nous utilisons également la segmentation de marché, basée sur une méthodologie mixant trois techniques : l'analyse par composante principale, la segmentation hiérarchique et les K-means. Chacune de ces trois façons de segmenter comporte des forces et faiblesses, nous avons donc souhaité tirer le meilleur des trois techniques, afin de développer notre propre technique de segmentation.

Nous segmentons, depuis un an et demi, les différents types de donateurs, ce qui nous a permis de revoir notre stratégie marketing sur chacun des segments. Nous avons une démarche plus ciblée et commençons le traitement de la Customer Journey, du parcours consommateur complètement basé sur les données. Quel est le segment ? Quelle est la proposition

de valeur faite à ce segment d'affaires ? Quel est le parcours client pour pouvoir livrer cette expérience ? Quels sont les endroits de l'entreprise où l'on fait cela bien, ou alors les endroits où le parcours est brisé ? On va avoir de plus en plus, un parcours client basé sur les données et non sur les intuitions du marketing.

Nous souhaitons enfin utiliser le Text Mining sur des données non structurées. Nous avons énormément de texte saisi dans nos centres d'appel. Il s'agit d'une grande opportunité pour améliorer l'expérience client. **Selon Oracle, 80% des données qui existent dans le monde sont non structurées.** Les données financières et transactionnelles le sont déjà. Naturellement, si j'ai un conseil à donner, c'est de commencer par les données structurées disponibles, mais ce sont des données « after the fact », par exemple le cas d'une annulation, d'une cessation de paiement ou un déménagement. Les données non structurées sont souvent des données qui reflètent ce qui se passe à un temps T. Ce sont des « Leading Indicators » qui vont être prédictifs et pas uniquement descriptifs.

### QUELLES SONT LES DIFFÉRENTES SOURCES DES DONNÉES RÉCOLTÉES ?

Nous collectons les données transactionnelles (dons, interactions entre donateurs et Vision Mondiale) dans une nouvelle base de données. **Nous avons souhaité construire une image globale du donateur, une vue 360° du donateur.** Nous avons ainsi construit un entrepôt de données, structuré autour du donateur. Les données électroniques sont récoltées d'une manière descriptive, mais nous voulons les ramener dans un endroit centralisé pour les marier avec les données transactionnelles. Sur la partie authentifiée de notre site web, nous pouvons aller chercher les données transactionnelles du donateur identifié pour un parcours personnalisé à son segment d'affaires. Nous utilisons également des données de recherche externes. Nous avons mis en place deux plateformes électroniques pour faire de la recherche : un panel sur la population canadienne et un sur les autres donateurs. Nous utilisons souvent ces panels pour tester les actions et campagnes marketing. Il y a 3 ans, nous faisons un ou deux sondages par an, maintenant nous réalisons une quarantaine d'études, cela nous permet de réagir très rapidement et d'être plus agile. Nous générons des rapports ou des études mais nous emmagasinons également les données dans une base de données pour voir l'évolution des réponses sur les trois dernières années. Ces données sont ensuite intégrées dans IBM Modeler. Nous n'utilisons pas uniquement ce que nos donateurs nous disent mais égale-

ment qui ils sont, nous essayons ainsi d'étudier leur comportement post étude. **Les données ne peuvent plus être uniquement transactionnelles, toutes les données doivent être rapprochées pour avoir une histoire globale.**

### EST-CE QUE VOUS AVEZ OBSERVÉ DES RÉSULTATS DIRECTS DE L'ANALYTIQUE MIS EN PLACE ?

Au niveau de l'utilisation de la segmentation et des modèles prédictifs, nous avons par exemple pu cibler nos efforts sur un des segments, qui sponsorise un enfant par donateur avec une moyenne de 30 euros par mois, en revanche ils ne donnaient rien d'autre et ne soutenaient aucun autre programme. Malgré nos tentatives de communication et mails personnalisés, nous n'avions pas de réponse. Nous avons pu les isoler, nous avons utilisé les « online panels », fait des focus groupes, des rencontres clients et nous nous sommes rendus compte qu'ils aimeraient communiquer avec l'enfant aidé et avoir une relation à travers des lettres par exemple. Nous avons fait un test en envoyant une enveloppe transparente indiquant qu'il s'agissait uniquement d'une communication pour aider à prendre contact avec l'enfant, sans demande de dons. **Grace aux insights data, nous sommes passés d'un taux de réponse de 0 à 20% : nous les avons fait réagir.**

### VOUS DITES POUVOIR VOUS INSPIRER DES ENTREPRISES PRIVÉES ? EST-CE QUE VOUS AVEZ DES EXEMPLES D'ENTREPRISES QUI ONT PU VOUS SERVIR D'INSPIRATION ?

Nous pouvons tirer inspiration du domaine de la **fraude bancaire**, nous devons apprendre de leurs techniques et manière d'utiliser les statistiques avancées, pour comprendre qui sont les influenceurs et les ambassadeurs pour les futurs donateurs. Au niveau du commerce électronique, nous pouvons également nous inspirer des **recommandations faites pour des produits complémentaires** sur Amazon par exemple, qui viennent ajouter de la valeur pour le donateur et non pour l'organisme uniquement. **Cela est rendu possible avec des données structurées et non-structurées et par une réaction rapide.** Nos donateurs passent en moyenne 5 à 7 minutes sur notre site web, il faut que nous puissions non seulement avoir la donnée (elle est générée très rapidement si elle est taguée) mais surtout avoir derrière un engin de calcul live pour nous permettre de faire des recommandations business.

**Notre objectif final est de marier toutes ces données récoltées sur le moment avec des données non structurées : offline, segments, modèles.**








Selon Oracle, **80%**  
**des données** qui existent  
dans le monde **sont non**  
**structurées**

# Entreprises










# & LABELS



	 Analytique	 API	 Applications	 Consultants/SSII Intégrateurs	 Dataviz	 Infrastructure	 Institutionnels
AB INITIO SOFTWARE						•	
ACCENTURE	•			•	•		
ADVANCED SCHEMA	•		•	•	•		
ALTARES	•	•	•	•	•	•	
AMAZON WEB SERVICES	•						
ATTUNITY				•			
BEARINGPOINT	•			•			
BLUE DME	•		•				
BLUESCALE		•	•	•	•		
BLUESOFT GROUP	•			•		•	
BMB SERVICES				•			
BUSINESS & DECISION				•			
CENTRALESUPÉLEC EXED							•
CISCO FRANCE	•	•	•			•	
CLOUDERA FRANCE	•		•			•	
CLUB DECISION DSI							•
COHERIS	•				•		
COMPLIANCE IT LEGAL CONSULTING				•			
CONVERTEO				•			
COUCHBASE						•	
COUTHON CONSEIL				•			
DATASTAX	•		•				
DENODO	•						
DIGDASH	•		•		•		
ECS - MEDIASCHOOL EXECUTIVE EDUCATION							•
EKIMETRICS	•	•	•	•	•		
ELASTIC	•	•	•		•		
EULIDIA	•			•	•	•	
EXALEAD DASSAULT SYSTEMES	•		•		•		
EXPERIS IT	•			•			
FINAXYS				•			
GFI INFORMATIQUE			•	•	•	•	
HEWLETT PACKARD ENTERPRISE	•	•		•		•	
HURENCE	•		•	•		•	

# BIG DATA

	 Analytique	 API	 Applications	 Consultants/SSII Intégrateurs	 Dataviz	 Infrastructure	 Institutionnels
IDEATRANS	•		•	•	•		
KALANE				•			
KEYRUS	•		•	•	•		
MAPR TECHNOLOGIES SAS	•		•			•	
MARKLOGIC FRANCE				•			
MATHWORKS	•		•		•		
MATLO	•		•		•		
MICROPOLE	•	•	•	•	•		
MONGO DB	•		•		•	•	
NEO TECHNOLOGY	•					•	
ONEPOINT				•			
OUTSCALE		•	•			•	
PENTAHO	•			•	•		
PROXEM	•	•	•		•		
QLIK	•				•		
REPORT ONE	•			•	•		
SAAGIE	•		•	•	•	•	
SAS INSTITUTE S.A.S	•				•		
SCALED RISK	•		•				
SCINETIK	•			•			
SEENK LAB		•	•	•	•		
SENTELIS				•			
SINEQUA	•		•				
SKAPÁNĚ	•		•	•			
SOFT COMPUTING				•			
SOPRA STERIA			•	•			
SPLUNK			•				
SYNALTIC GROUP	•		•	•	•		
TABLEAU FRANCE					•		
TALEND						•	
TELECOM PARISTECH / EVOLUTION							•
TERADATA	•					•	
TIBCO SOFTWARE	•	•	•		•	•	
YSANCE	•			•	•		

+33(0)1 42 34 90 00  
 solutions@abinitio.com  
 1 rue Danton 75006 Paris France



www.abinitio.com



Les logiciels Ab Initio constituent une plate-forme universelle de traitement et de gestion de données d'entreprise. Cette architecture unique permet le traitement de fichiers, de tables de bases de données, de files de messages, de services Web et de métadonnées. Cette architecture permet de définir, partager et exécuter n'importe quelle règle technique et métier au travers d'un outil graphique. Elle traite les données en parallèle sur plusieurs processeurs, éventuellement distribués sur des serveurs différents. Les mêmes règles peuvent être exécutées en mode batch ou temps réel, et dans une architecture orientée service. Les fonctions d'alerte et de supervision permettent un contrôle fin des applications en production et les traitements en erreur peuvent être redémarrés rapidement à partir des points de reprise intégrés nativement. Cette même architecture permet à des utilisateurs non techniques de collecter, versionner et analyser les métadonnées de bout en bout.

**Les plus grandes entreprises mondiales ont choisi Ab Initio pour les applications clés de leur entreprise. Voici quelques exemples :**

- **Cartes de Crédit :** Un réseau de cartes de crédit utilise la solution Ab Initio pour son infrastructure de gestion des données, ce qui lui permet de traiter l'ensemble des transactions et de les transmettre en mode batch ou en temps réel aux différentes applications du système d'information. De nombreuses applications en aval, notamment la facturation commerciale, sont également implémentées avec Ab Initio. Un pétaoctet de données de transactions est stocké chaque année dans un système d'archivage Ab Initio, permettant au service clientèle de répondre instantanément aux diverses requêtes des clients.
- **Assurance :** Un des principaux acteurs du marché de l'assurance utilise Ab Initio pour de nombreux aspects du traitement des demandes d'indemnisation. Le système de traitement des contrats de réassurance en place dans cette société comporte des milliers de règles de gestion complexes qui sont toutes mises en œuvre avec Ab Initio.
- **Banque :** Une banque internationale utilise la solution Ab Initio pour consolider les informations relatives à l'ensemble des clients, tous secteurs confondus, dans un vaste entrepôt de données. Ab Initio lui permet également de personnaliser et de traiter tous les flux de transactions SWIFT entre ses filiales internationales.
- **Bourse :** Une grande place boursière a remplacé des millions de lignes de code Cobol par des applications Ab Initio destinées à des opérations vitales pour l'entreprise. Connectée au système temps réel de trading, elle permet de traiter les transactions au rythme de 500 000 messages par seconde.
- **Distribution :** Un grand distributeur contrôle son inventaire et peut détecter d'éventuelles fraudes en recevant des données en temps réel de ses terminaux points de vente provenant de plusieurs milliers de magasins.
- **Telecoms :** Un des plus grands opérateurs téléphoniques mondiaux utilise Ab Initio pour traiter les informations de comptes rendus d'appels dans le cadre du calcul des coûts d'appel, du suivi d'utilisation et du contrôle du réseau. Plusieurs milliards d'enregistrements de comptes rendus d'appels sont traités chaque jour ainsi que des millions de demandes relatives à l'usage du réseau.
- **Internet :** Une des plus grandes sociétés Internet traite chaque jour des dizaines de milliards de clics sur les annonces en ligne dans le cadre de ses services de facturation et d'optimisation des placements publicitaires.

**COMMENT EXPLIQUER QUE TOUTES CES ENTREPRISES SE SOIENT Tournées vers AB INITIO ?**

Les produits Ab Initio sont intuitifs et simples d'utilisation. Ils sont également capables de gérer les logiques applicatives les plus complexes ainsi que des volumes de données considérables. Cela avec des performances exceptionnelles et une fiabilité remarquable. Cette combinaison est tout simplement unique.

Depuis 20 ans, Ab Initio se concentre sur le traitement et la gestion des données d'entreprise, y compris les très gros volumes (de centaines de téraoctets à plusieurs pétaoctets), les logiques applicatives les plus complexes, en batch et en temps réel, et la gestion complète des métadonnées. Bon nombre des plus grandes entreprises mondiales ont logiquement choisi Ab Initio pour résoudre les problèmes très complexes liés à la gestion de leurs données. C'est dans ce contexte qu'Ab Initio propose une solution complète pour lire, écrire, traiter et interroger les «big data» stockées dans Hadoop. Cette solution utilise toute la puissance qui fait la réputation d'Ab Initio, notamment :

- Une suite logicielle totalement intégrée
- Une plate-forme de développement graphique hautement productive et facile à gérer
- La possibilité d'exécuter des applications en parallèle, parfaitement évolutives, sur des clusters hétérogènes regroupant l'ensemble des principales plates-formes (Unix, Linux, Windows, Mainframe)
- Performances élevées pour le traitement en mode batch et en temps réel
- Prise en charge de l'intégration avec les métadonnées de l'entreprise via la solution de métadonnées d'Ab Initio
- Interrogation fédérée des données réparties sur l'ensemble de l'entreprise : dans un système HDFS ou dans des bases de données relationnelles, des fichiers plats, des datasets SAS, des feuilles de calcul Excel, entre autres.

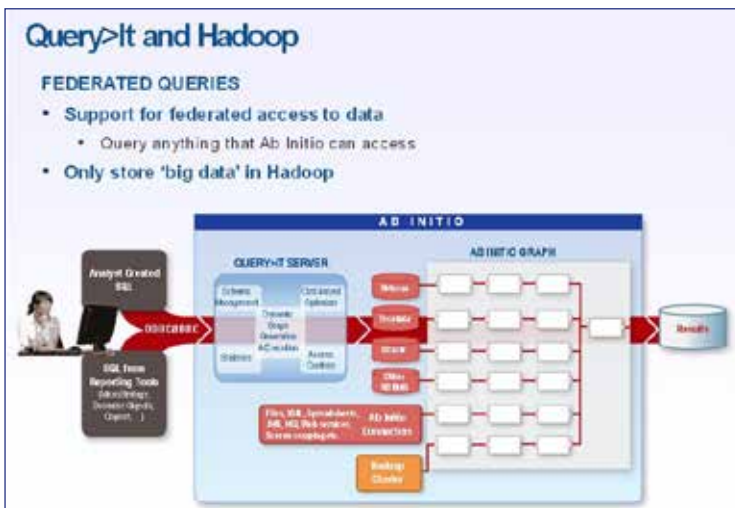
Le degré de parallélisme d'une application Ab Initio peut être ajusté en fonction de la répartition des données sur le cluster Hadoop. Ab Initio s'intègre aux séquences et aux fichiers texte Hadoop et prend en charge les formats Parquet, RCFile et ORC de base. Les applications Ab Initio peuvent également être intégrées à Yarn.

### Requêtes Hadoop

Ab Initio propose des composants intégrés permettant la lecture et l'écriture des données Hadoop en parallèle. Les applications Ab Initio assurent le traitement simultané des données Hadoop et de tout autre type de données (SGBD, fichiers plats, flux XML, file de messages,...) sur différents types de plates-formes (Linux, AIX, Windows, z/OS, par exemple).

Ab Initio permet d'appliquer des mécanismes de sécurité aux sources de données Hadoop au niveau des colonnes et des lignes, afin de garantir que seuls les utilisateurs et les groupes disposant des privilèges adéquats puissent y accéder.

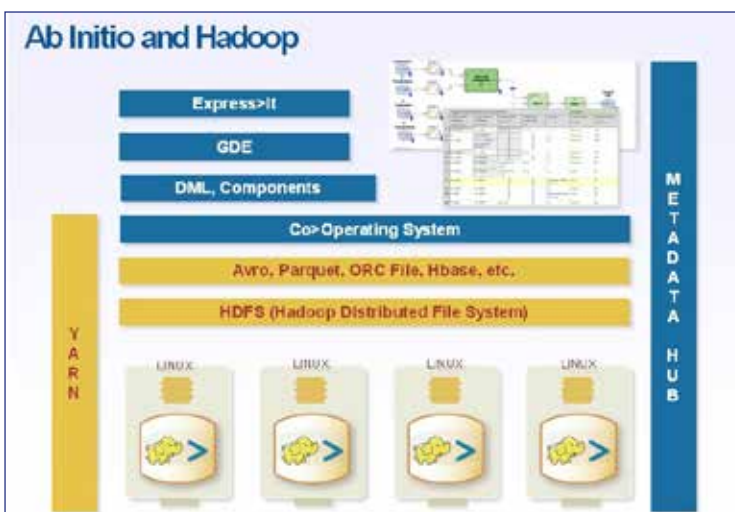
Ab Initio permet l'accès indexé aux fichiers Hadoop, beaucoup plus rapide et efficace que les technologies de type Hive.



### Requêtes fédérées et création de datasets Hadoop

Ab Initio propose un accès fédéré aux données, ce qui permet aux utilisateurs d'interroger toutes les bases de données ou tous les fichiers, y compris Hadoop, auxquels Ab Initio peut accéder.

Lorsque des utilisateurs soumettent des requêtes SQL fédérées via Ab Initio, ils peuvent charger les résultats dans une cible prise en charge par Ab Initio, y compris Hadoop. Cette fonctionnalité permet aux analystes et aux autres utilisateurs de créer et d'exploiter des datasets Hadoop ad hoc à partir de données provenant de nombreuses autres sources.





## LES ARCHITECTURES BIG DATA HYBRIDES POUR ACCÉLÉRER « LA LOGISTIQUE DE LA DONNÉE »

Au cours des dernières années, nous avons été les témoins d'une gigantesque transformation dans la gestion et l'utilisation de la donnée. Les solutions ont évolué de simples entrepôts de données (Data Warehouse) à des architectures Big Data. Les premiers usages ont oscillé entre le « Datalake » où les données sont stockées, souvent dans une perspective Analytics, et les approches « opérationnelles » où le Big Data vient optimiser des processus existants (réduction des coûts DWH, mise en cache du mainframe, ELT distribué...).

Aujourd'hui ces approches tendent à converger, introduisant le besoin d'architectures hybrides dans lequel nous retrouvons l'écosystème Hadoop, mais aussi des entrepôts de données (EDW) massivement parallèles (MPP), des bases de données orientées colonne « in-memory », Stream Computing, NoSQL, voire d'autres approches supportant l'Extreme Analytics dans le Cloud.

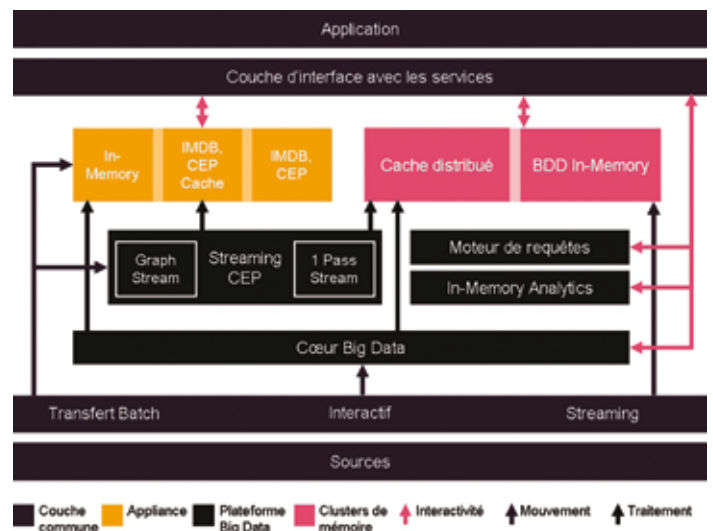
Ces architectures hybrides adressent la réalité hétérogène des environnements Big Data et répondent au besoin d'incorporer de nouvelles comme d'anciennes bases de données analytiques dans une architecture commune « centrée sur la donnée ».

Pour mettre en musique les différents composants de ces architectures hybrides, il faut concevoir une vraie « logistique de la donnée », et introduire le concept de « donnée accélérée » qui puise ses vertus dans des outils et technologies qui permettent d'ingérer de très grands volumes de données. Il permet également aux organisations d'accéder instantanément à des données clés pour une analyse et une prise de décisions dans des délais très courts, ce qui constitue un avantage majeur sur un marché concurrentiel.

### La donnée accélérée introduit 3 challenges :

- **Mouvement** : Comment faire transiter la donnée de sa source jusqu'à l'emplacement cible de l'architecture ?
- **Traitement** : Comment la traiter pour en extraire la valeur ajoutée, actionnable très rapidement ?
- **Interactivité** : Comment accélérer l'accès à la donnée suite aux requêtes des utilisateurs ou applications ?

Bien sûr, tous les contextes ne justifient pas l'ensemble de ces briques, et la difficulté réside dans le choix des types de composants, puis la solution associée à chacun, et enfin la mise en musique globale. Une approche « technique » part des différents modèles d'architecture qui ont cours (Lambda, Kappa, SMACK...), mais le marché peine encore à s'accorder sur ces concepts qui donnent lieu à de vrais débats d'expert difficiles à concilier avec une approche métier.



Accenture a donc développé une méthodologie adaptée à ces enjeux : la « **Data Supply Chain Diagnostic Methodology** ».

Cette méthodologie repart des problématiques métier, et s'appuie sur des accélérateurs (questionnaires, comparatifs de solution, retours d'expérience en production...) pour définir l'architecture cible et sa déclinaison technique.

Avec cette méthodologie Accenture accompagne aujourd'hui ses clients pour accélérer la logistique de la donnée.

### Exemples d'approches selon les types de besoins

	Mouvement « Se brancher au tuyau, et ne pas perdre une seule goutte »	Traitement « Besoin de traiter de grands volumes de données plus rapidement »	Interactivité « Questions urgentes nécessitant des réponses rapides »
	ETL	Streaming	
1. Appliance seule	Basique	Avancé	Avancé
2. BDP vers Appliance	Basique	Avancé	Avancé
3. Streaming vers Appliance	Basique	Avancé+	Avancé
4. BDP seule	Basique	Basique	Basique
5. Streaming vers BDP	Basique	Avancé	Basique
6. BDP avec Analytics In-Memory	Basique	Avancé	Avancé
7. Streaming vers BDP avec Analytics In-Memory	Basique	Avancé+	Avancé
8. BDP avec moteur de requêtes	Basique	Basique	Avancé
9. Streaming vers BDP avec moteur de requêtes	Basique	Avancé+	Avancé
10. Cluster de cache distribué	Basique	Basique	Avancé
11. BDP vers Cluster de cache	Basique	Basique	Avancé
12. Cluster de base de données In-Memory	Basique	Basique	Avancé
13. BDP vers Cluster de base de données In-Memory	Basique	Basique	Avancé
14. Streaming vers Cluster de base de données In-Memory	Basique	Avancé+	Avancé
	Le traitement d'évènement complexe peut améliorer l'ingestion en streaming	Le traitement d'évènement complexe peut augmenter la vitesse en prétraitant les données	Les bases de données Cache et In-Memory permettent l'interactivité en temps réel

# Pierre-Yves Lesage

Responsable Analytics Delivery en France

## INTERVIEW



Accenture a inauguré cette année, à Paris, un centre d'innovation pour aider les entreprises à imaginer les innovations de demain, anticiper leurs futurs modèles économiques, repenser leur stratégie numérique et créer de nouveaux produits et services. Ce nouvel environnement immersif dédié à l'innovation propose aux entreprises des parcours d'innovation personnalisés, de la génération d'idées jusqu'à leur concrétisation. Ils sont construits par des équipes multidisciplinaires d'Accenture autour d'une problématique business et d'objectifs spécifiques à l'entreprise participante. Au cœur de ces parcours immersifs, l'offre **Accenture Connected Analytics Experience** a été développée par Accenture Labs, structure dédiée à la recherche et au développement technologique (R&D), pour infuser l'Analytics dans les processus innovants.

### CONCRÈTEMENT, QU'EST-CE QUE L'ACCENTURE CONNECTED ANALYTICS EXPERIENCE ?

Il s'agit d'une expérience à vivre autour de l'Analytics, immergé dans « l'Analytics room » une salle et un environnement dédiés. Cette expérience a été conçue pour remplir deux objectifs : rendre l'Analytics plus facile à comprendre et permettre aux utilisateurs métiers d'entreprendre un parcours intégré : d'un C-level à un manager fonctionnel, et réunir les acteurs de la données, de l'utilisateur final au data scientist, afin de construire une solution « actionnable », c'est-à-dire directement intégrable au sein des processus métier de l'entreprise. L'idée est de pouvoir faire collaborer différents acteurs de l'entreprise, réunis autour d'une même problématique dans un environnement dédié, en analysant de manière interactive et restituant des informations sous forme visuelle, rendant les données plus lisibles et compréhensibles.

### EN QUOI CETTE EXPÉRIENCE EST-ELLE PARTICULIÈRE ?

La collaboration est l'enjeu primordial dans cette approche. Aujourd'hui nous constatons que de nombreux projets Analytics ont du mal à passer d'une phase de « Proof of Concept » à une phase d'industrialisation qui applique l'Analytics au sein des processus métiers. L'expérience montre que les POC sont souvent proposés et lancés par les équipes techniques ou DataScientists, mais peu avec une implication suffisante des métiers. Accenture a souhaité proposer une approche différente, qui valorise la collaboration entre les acteurs de l'organisation, les données et la technologie, pour donner du sens, partager une vision, co-construire et finalement fédérer autour d'objectifs business communs.

Cette offre est articulée autour de 5 piliers : **Stratégie Analytics** : la définition de la vision et de la stratégie autour de la donnée, tant sur des aspects d'organisation, de gouvernance que de mise en place de centres de compétences Analytics ; **Experts Données** : des équipes pluridisciplinaires avec des expertises variées comme l'ingénierie des données, la data science, la visualisation... ; **Centre d'Excellence Analytics** : des centres d'expertise spécialisés sur des problématiques Analytics pointues telles que Fraud Analytics (Dublin, Irlande), Operation Analytics (Barcelone, Espagne) et Customer & Marketing Analytics (Athènes, Grèce) ; **Plateforme Données** : une plateforme Big Data, déployée dans un Datacenter Accenture en France et prête à l'emploi pour implémenter et accélérer la découverte d'insights. Cette plateforme est directement connectée et intégrée à l'Analytics Room ; **Environnement immersif** : toutes ces composantes sont assemblées au sein de l'Analytics Room, par l'intermédiaire de la technologie Mezzanine de l'éditeur Oblong, qui permet de concentrer les acteurs dans un environnement innovant, collaboratif et immersif.

### QUELS SONT LES AVANTAGES DE L'ACCENTURE CONNECTED ANALYTICS EXPERIENCE ?

Il s'agit d'une offre complète, au service des entreprises ayant pour ambition de construire et d'inspirer à leurs équipes une culture de l'Analytics, permettant ainsi de construire un agenda solide sur ce sujet. L'environnement immersif permet une véritable collaboration de toutes les parties impliquées dans la prise de décision et tout cela, en temps réel. C'est souvent le point de départ d'un projet d'ampleur qui démarre directement avec un sponsor métier.

### A QUEL TYPE D'ENTREPRISE S'ADRESSE CETTE OFFRE ?

Les entreprises dans les télécommunications, le service de la santé, les biens de consommation, les industries et les services financiers font déjà parti de l'Accenture Analytics Connected Experience. L'Analytics peut aider les entreprises dans toutes les industries, mais aussi les gouvernements à prendre des bonnes décisions pour atteindre leurs objectifs, que ce soit l'amélioration de l'expérience client, la santé ou la sécurité publique. A titre d'exemple, Accenture travaille actuellement avec un fournisseur d'eau potable au Royaume-Uni pour l'aider à anticiper les défaillances de l'équipement et répondre en temps réel à des situations critiques comme les fuites ou phénomènes météorologiques défavorables. Accenture aide également une agence gouvernementale à analyser le suivi de flux vidéo afin d'identifier les risques potentiels pour la sécurité publique. Toutes les entreprises disposent actuellement d'une grande quantité de données et d'outils d'analyse. Le point essentiel est désormais de savoir collaborer pour mieux les exploiter afin de permettre aux différents services de prendre des décisions précises et rapides sur la base des données et des résultats d'analyse. L'entreprise qui saura tirer profit de l'Analytics en s'appuyant sur la collaboration, sera alors dans la meilleure position pour défendre, faire évoluer et transformer ses marchés.



## ADVANCED SCHEMA

**HERVÉ SEBAN**

Président

+33 (0)6 03 22 89 15

hseban@advanced-schema.com

19, rue Galilée - 75016 Paris



Analytique



API



Applications



Consultants/SSII  
Intégrateurs



Dataviz



Infrastructure



Institutionnels

www.advanced-schema.com



**advanced**  
schema

**ADVANCED Schema** est une Entreprise internationale de Services du Numérique (ESN), présente en Europe et en Amérique du Nord, qui intervient sur 5 domaines d'activités stratégiques : la Business Intelligence, le Data Warehousing, le Big Data, le CRM (ou Gestion de la Relation Client) et le Digital.

Ainsi, et depuis près de 15 ans, ADVANCED Schema accompagne et conseille ses clients dans leurs projets de refonte de leurs systèmes d'information et dans la mise en œuvre de systèmes de pilotage et de reporting performants.

Dans un univers toujours plus digitalisé où la donnée est au cœur des problématiques métiers, le Groupe ADVANCED Schema s'est structuré afin de pouvoir proposer à ses clients des offres adaptées en lien avec ces nouveaux enjeux.

Une organisation nouvelle illustrée notamment par la création d'un pôle R&D en 2011 et par l'acquisition de la société Kernel42 spécialisée dans le Digital en 2014.

Afin de proposer une offre qui soit la plus souple possible, notre équipe R&D décline son savoir-faire à travers deux offres distinctes :

- le « Lab as a Service », véritable plateforme répondant aux besoins d'expérimentation et de flexibilité de ses clients,
- le mode « Tailored » pour un développement à façon d'applications décisionnelles et Big Data.

Dernière-née du groupe, la société nodata, spécialisée dans l'édition de logiciels, présente NODATA une solution intégrée d'outils autour de la donnée. Alliant modularité, performance et maîtrise des coûts, la suite NODATA est le fruit d'un mariage réussi entre le savoir-faire historique d'ADVANCED Schema et des compétences nouvelles développées au sein de son pôle R&D.



## FOCUS SUR LES FILIALES DU GROUPE ADVANCED SCHEMA

**nodata**

**Réinventez l'exploitation de la donnée** avec **nodata**, nouvelle filiale du groupe entièrement dédiée à l'édition de logiciel. La **suite Nodata (Not Only Data)** offre la possibilité d'allier innovation technologique, performance et expérience utilisateur pour des coûts maîtrisés.

Cette solution complète et totalement modulable couvre - par exemple - aussi bien la transformation de la donnée, la modélisation, la data visualisation...



**Kernel 42**

**Kernel 42** étend d'expertise d'ADVANCED Schema aux projets digitaux : **Web, e-commerce, mobilité et Big Data.**

La combinaison de ses expertises digitales et Data (Big Data, Business Intelligence) permet de concevoir des **dispositifs digitaux intelligents** et parfaitement intégrés au Système d'Information.



ADVANCED Schema lance Connexin, sa filiale **spécialisée dans la santé connectée**. Connexin est une plateforme de coordination et de communication entre professionnels de santé.

Elle a été développée afin de **faciliter la communication dans le lien ville-hôpital** entre professionnels de santé en utilisant les **canaux digitaux**.



# Nicolas Brigitte-Alphonsine

Directeur Général

## INTERVIEW



### L'AN DERNIER À LA MÊME ÉPOQUE VOUS NOUS PRÉSENTIEZ LA SOLUTION NODATA TOUT DROIT SORTIE DES LABS R&D DU GROUPE ADVANCED SCHEMA. POUVEZ-VOUS NOUS DIRE OÙ VOUS EN ÊTES UN AN APRÈS ?

Depuis le dernier salon Big Data où notre solution Nodata a suscité un vif intérêt nous avons énormément travaillé et nous sommes fiers de compter aujourd'hui 5 références clients grands comptes parmi lesquels 3 ambassadeurs de la solution. Ces ambassadeurs contribuent à éprouver la solution Nodata sur des problématiques complexes qui dépassent le simple cadre d'un PoC (Proof of Concept). Nodata répond aux contraintes d'un déploiement industriel et respecte une philosophie *release early/release often* nous permettant ainsi de proposer des améliorations régulières et constantes à nos clients. La solution Nodata, s'illustre au travers de cas d'utilisation : Améliorer le pilotage de l'entreprise, faciliter l'accès à la donnée, mais aussi, accélérer le *time to market* des projets.

Notre équipe R&D – qui n'a cessé de grandir et qui compte aujourd'hui plus de 30 développeurs – va poursuivre ses efforts afin d'enrichir la solution Nodata.

### QUELLE EST LE RÔLE DE L'ÉQUIPE R&D SUR LES TECHNOLOGIES BIG DATA ?

Les technologies dites Big Data sont en ébullitions et sont le fruit d'initiatives tous azimuts de la part des éditeurs, des universités et des individus. Le degré de maturité de ces technologies est très variable, le niveau de modularité des composants est très élevé, et la diversité des langages et des briques sous-jacentes est très importante.

Dans un monde de la donnée où les technologies prépondérantes permettaient un accès aux données en SQL, l'approche NoSQL traitée par micro-batching ou streaming nécessite une maîtrise importante de langages telles que Java ou Python par exemple. L'interconnexion de ces différents composants se fait principalement via une sérialisation en JSON au-dessus du protocole HTTP, souvent avec une approche RESTful.

Les compétences concentrées dans notre équipe R&D nous ont permis d'adresser ces technologies de manière efficace, en ayant la capacité d'évaluer la maturité et la performance de celles-ci au travers de cas d'usage concrets amenés par plus d'une décennie d'expériences autour de la *data*.

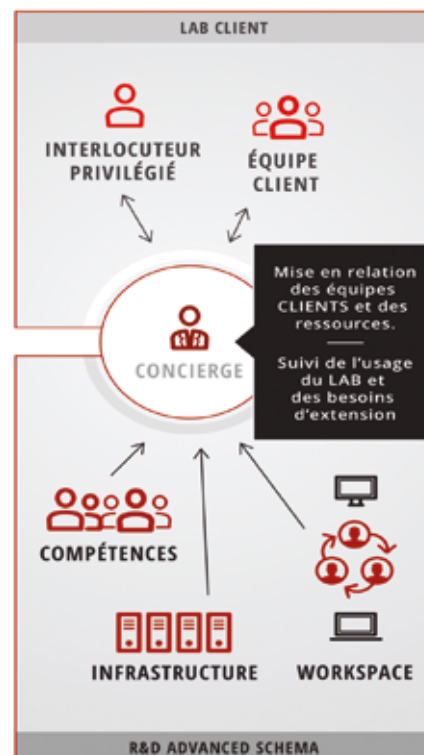
Fort de ces capacités nous avons pu identifier, expérimenter, et mettre en œuvre des technologies aujourd'hui phares telles qu'ElasticSearch, Spark, HBase... La montée en compétence des équipes projets est un point fondamental pour l'ensemble de nos clients. Il est donc primordial d'être en mesure de les accompagner en nous appuyant sur nos expertises technologiques et nos retours d'expériences.

### QUELLES SONT LES OFFRES D'ADVANCED SCHEMA QUI PERMETTENT D'ACCOMPAGNER VOS CLIENTS SUR LE VIRAGE DU BIG DATA ?

Un grand nombre de nos clients ont émis le souhait de monter en compétence sur des technologies de pointes et désirent éprouver la valeur ajoutée apportée par ces dernières. Pour réussir cette montée en compétence sur des technologies Big Data, il est indispensable d'allier maîtrise des coûts et synergie (en terme d'équipe, de méthodologies et de technologies).

En réponse à ce défi, nous avons décidé de créer une offre que nous avons baptisé **Lab as a Service**. Cette offre s'articule autour d'un **Concierge** qui est mis à la disposition de nos clients et qui propose des services spécifiquement adaptés à leurs besoins métier et IT. Avec le **Lab as a Service**, nous apportons nos compétences techniques, notre infrastructure à travers notre Cloud privé et nous accompagnons intégralement la réalisation et le suivi des expérimentations de nos clients.

À l'image de l'ensemble des services *on demand*, cette offre permet une très grande souplesse de facturation, pour une parfaite optimisation des coûts.



**Yohan WASMES BENQUE**

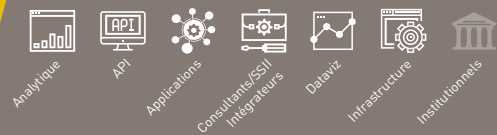
Chief Marketing and Innovation Officer

Twitter : @yohanbenque

+33(0)6 50 70 29 54

Yohan.benque@altares.com

Immeuble Le Capitole - 55 - av des Champs Pierreux  
92012 Nanterre France



www.altares.com



**SCOPFI BY ALTARES : COMMENT LE LEADER DE LA DATA D'ENTREPRISE A DÉCIDÉ DE SE POSITIONNER COMME INTERMÉDIAIRE DE CONFIANCE ENTRE LES TPE/PME VOULANT SE FINANCER ET LES ORGANISMES BANCAIRES OU ALTERNATIFS GRÂCE À UNE PLATEFORME BASÉE SUR LA SMART DATA.**

La plateforme Scopfi by Altares a une double vocation : elle est d'une part un facilitateur de financement pour les TPE/PME et d'autre part un outil d'aide à la décision pour les organismes qui financent les entreprises. L'idée est de fonctionner comme un guichet unique qui permet aux entreprises d'optimiser leur chance d'obtenir la solution de financement la plus adaptée à leurs projets, en toute simplicité, rapidement, au meilleur coût, et 100% online. Via la plateforme, elles peuvent connaître en temps réel leur capacité de financement, elles peuvent disposer d'un dossier pré-rempli à envoyer aux établissements de financement et elles bénéficient d'une aide à la décision objective et rapide.

Pour les organismes de financement (banques, crowdfunders, sociétés d'affacturage) cette plateforme est un nouveau canal digital de commercialisation qui leur fait bénéficier de dossiers prospects qualifiés. Elle allège aussi le travail d'analyse des chargés d'affaires grâce à la mise à disposition d'une évaluation de la qualité des demandeurs, elle facilite la sélection des entreprises en fonction de leurs critères d'éligibilité, elle maîtrise leur canal de conversion en benchmarkant la concurrence, en testant et en pilotant leurs offres en quasi temps réel : type de financement, critère d'éligibilité, niveaux de garantie et enfin elle met à disposition des tableaux de bord (statistiques de financements): nombre de demandes, nombre d'offres, nombre de signatures, etc. L'idée de rapprocher financeurs et financés n'a rien de nouveau, les enjeux sont connus et de nombreux acteurs, publics et privés, en connaissent l'importance. Même si son activité principale n'est pas le financement d'entreprise, Altares a souhaité accompagner ses clients jusqu'au bout de leur démarche grâce à l'utilisation du digital, du Smart Data et d'algorithmes décisionnaires.

Par ailleurs, le sujet du financement n'est pas nouveau pour Altares. Avec ses études trimestrielles sur les défaillances d'entreprises, l'entreprise sait parfaitement bien que le manque de financement à un moment clef du développement des sociétés est une des premières causes de défaillance. Partenaire de la Médiation des Entreprises depuis plusieurs années, Altares met à disposition son baromètre et ses analyses pour fluidifier le crédit inter-entreprises.

La mise en œuvre du projet s'est faite selon une méthode maîtrisée par Altares, celle très itérative du Lean Startup et du Pitch Elevator. Le « Go to Dev » a été lancé en juin 2015, puis se sont enchaînées les étapes de Product Discovery puis de Product Design, le BP, les études de marché, etc. Le salon Banque et Innovation du 1er octobre 2015 a initié la 2ème grande étape qui fut celle du lancement des partenariats et la réalisation du produit. Même si le développement a commencé depuis l'été, la décision a été prise d'accélérer les étapes de ce dernier. L'intérêt et l'adhésion des financeurs (crowdlenders, affactureurs, ou banquiers) a rassuré sur le potentiel marché et la vraie valeur intrinsèque et différenciante du produit.

**Le retour terrain a permis de designer au mieux le produit :**

- Intégration des spécificités liées à la typologie des financeurs (Crowdlender/Banque/Affacturage)
- Intégration des spécificités liées au métier des financeurs (Crédit vs mobilisation de créances)

Ce qui est intéressant, c'est que le modèle économique s'est précisé au contact du terrain et a découlé de la volonté de mettre en place un parcours client digital innovant.

De fait, La plateforme sera livrée dans les délais avec 10 offres de financeurs en ligne.

Ce positionnement digital et orienté clients oblige à aller beaucoup plus loin que de simplement détenir des datas, aussi complètes et riches soient-elles.

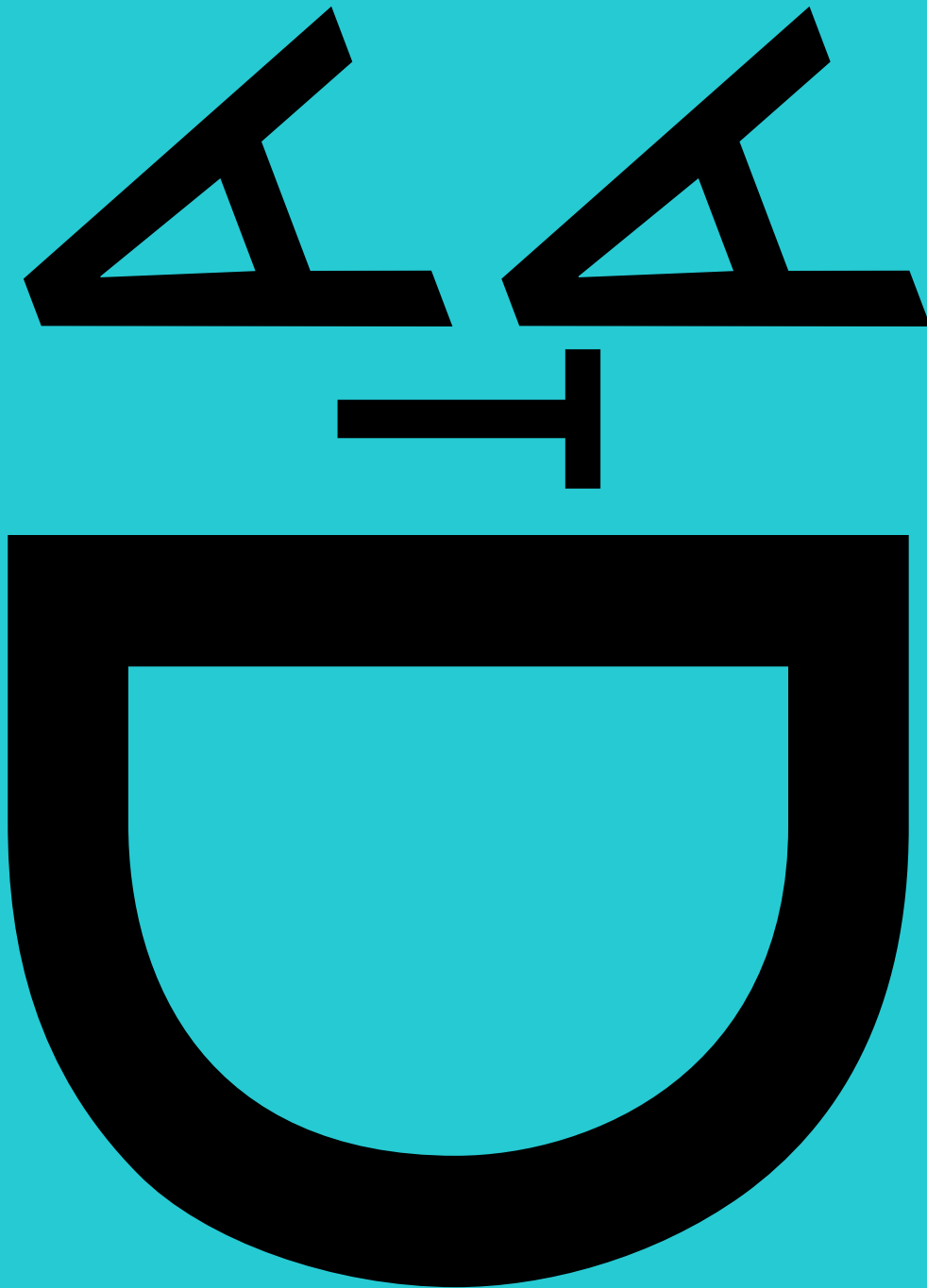
Il a donc fallu mettre au point plusieurs algorithmes permettant de mettre en relation les futurs « financés » de façon rapide et affine.

Le premier visait à filtrer les solutions de financement en prenant en compte à la fois le besoin exprimé et les caractéristiques de l'entreprise. Le second visait à donner un score de pertinence aux solutions de financement choisies : le montant du financement, son coût, son délai d'obtention et le caractère engageant de la solution choisie.

Enfin, il fallait également évaluer le profil du demandeur : notation du dirigeant principal, score de défaillance (données Altares), évolution des délais de paiement, dynamisme du secteur d'activité, etc.

Grâce à cette plateforme, Altares est aujourd'hui positionné sur un marché en fort développement qui englobe, de plus, des entreprises qui hésitaient jusqu'à maintenant à faire une demande de financement (2% des 3,2 millions de TPE/PME en France, soit 64 000 entreprises) et même les chefs d'entreprise pour qui le financement n'est pas un point de douleur seront sensibles à une démarche de simplification et d'efficacité comme celle que propose Altares.





Donnons du sens à la data...



CONTACT

# AMAZON WEB SERVICES

**Stephanie LAVOCAT**  
Head of France Marketing

PO BOX 84023 Seattle WA



# Leboncoin



Interview d'**Aïssa Belaïd**,  
Chief Data Officer -  
Leboncoin

«Dans le Cloud AWS, nous pouvons démontrer la viabilité d'un projet pour quelques centaines d'euros, contre plusieurs centaines de milliers d'euros avec notre infrastructure traditionnelle. (...) Essayer n'est plus un problème. Notre équipe est libérée et ose innover !»

### A PROPOS DE LEBONCOIN

leboncoin est la première plateforme généraliste de petites annonces en France. Plus d'un français sur trois a déjà utilisé son site fréquenté par 25,5 millions de visiteurs uniques par mois. Chaque jour, 800 000 à 1 million d'annonces sont déposées sur leboncoin.fr qui compte plus de 26 millions d'annonces en stock. Via sa régie publicitaire, une des dix premières en France, leboncoin propose aussi des services pour tous les annonceurs professionnels locaux et nationaux qui communiquent sur son site. Avec 400 employés et un chiffre d'affaires de 180 millions d'euros en 2015, leboncoin fait figure de site phare de l'économie collaborative et de la consommation raisonnée.

### LE DÉFI

leboncoin stocke et traite plus de 100 téraoctets de données issues de l'activité sur son site. Pour exploiter ces données, son équipe travaillait par le passé avec des outils traditionnels de Business Intelligence (BI) aux capacités d'analyse limitées. «La BI nous permettait d'analyser l'activité avec les chiffres des visites et des ventes réalisées sur deux semaines tout au plus. Sur ce laps de temps, les conclusions n'étaient pas représentatives et ne permettaient pas d'anticipation» explique Aissa Belaid, directeur de l'activité data leboncoin. Pour amener l'analyse des données au niveau de la prédiction, l'équipe devait s'assurer d'avoir une infrastructure capable de gérer des calculs massifs. Or son infrastructure hébergée dans deux data centers en colocation n'offrait pas la souplesse requise pour une approche big data. L'équipe data devait en continu anticiper les besoins en serveurs. «Gérer du stock de serveurs n'est pas notre métier et nous cherchions une solution plus commode, sans aucune limite de stockage» ajoute Aissa Belaid. Au plan budgétaire ensuite, l'infrastructure traditionnelle impliquant des investissements lourds à chaque nouveau projet. «Par exemple, créer notre plateforme de gestion des données (DMP) impliquait d'ajouter une cinquantaine de serveurs. Avec le rackage, l'installation, la location d'espace, le total s'élevait à près d'un million d'euros, ce qui était rétrograde dans notre infrastructure traditionnelle» explique Aissa Belaid. L'infrastructure dédiée aux big data devait évoluer et permettre de lancer de plus nombreux POC avec davantage d'agilité, sans de telles contraintes financières.

### POURQUOI AMAZON WEB SERVICES

Pour Aissa Belaid, l'infrastructure dédiée aux big data devait migrer dans le cloud pour échapper aux limites de stockage des data centers de l'entreprise. Le choix d'Amazon Web Services s'est rapidement imposé : «AWS offre la panoplie de services la plus riche du marché avec une grande stabilité que n'ont pas encore atteint les concurrents» souligne Aissa Belaid.

A présent, leboncoin utilise plusieurs centaines d'instances Amazon Elastic Compute Cloud (Amazon EC2) et peut en ouvrir à la demande le temps de créer et tester des POC. Son équipe utilise aussi le service de calcul serverless, AWS Lambda.

Son lac de données est stocké sur Amazon Simple Storage Service (Amazon S3). Avec Amazon Redshift, leboncoin dispose d'un service d'entrepôt de données rapide, entièrement géré et doté d'une capacité de plusieurs pétaoctets, auquel se combine le service Amazon Relational Database Service (Amazon RDS).

### LES AVANTAGES

Grâce aux capacités offertes par AWS, l'équipe data s'est affranchie de toute limite de stockage et sa gestion de projet bénéficie d'une plus grande souplesse. leboncoin stocke aujourd'hui plus de 100 téraoctets de données sur la plateforme d'AWS sans inquiétude quant aux volumes à venir. De plus, l'équipe n'est plus mobilisée par les questions complexes d'anticipation et de gestion des stocks de serveurs que soulevait son infrastructure traditionnelle. En utilisant AWS, elle dispose de ressources de stockage à la demande et peut ouvrir des instances Amazon EC2 sur mesure pour ses besoins d'expérimentation. «Nous avons repris la main sur notre métier !» apprécie Aissa Belaid.

Le modèle de facturation à l'usage d'AWS évite les investissements lourds en entrée, ce qui favorise le lancement de POC et encourage l'innovation. Aissa Belaid confirme : «Dans le Cloud AWS, nous pouvons démontrer la viabilité d'un projet pour quelques centaines d'euros, contre plusieurs centaines de milliers d'euros avec notre infrastructure traditionnelle. La prise de risque portant sur des sommes minimes, essayer n'est plus un problème. Notre équipe est libérée et ose innover !». En témoigne le projet de DMP en cours de développement sur la plateforme d'AWS : «Nous avons lancé notre projet de plateforme de gestion de données dans le Cloud AWS pour quelques milliers d'euros, alors qu'il était impossible de développer cet outil innovant dans notre infrastructure traditionnelle avec un coût de près d'un million d'euros en entrée».

Aujourd'hui, leboncoin s'appuie sur le Cloud AWS pour ses analyses big data. Alors que les anciens outils de BI se limitaient à une analyse de données sur deux semaines, leboncoin peut à présent exploiter les données de plusieurs années d'activité. «Dans le Cloud AWS, nous travaillons avec des data sets de 40 téraoctets de données sans aucun problème. D'ailleurs, nous ne nous occupons plus ni des coûts ni du stockage dans nos projets big data» indique Aissa Belaid. A la clé, son équipe peut identifier des tendances pertinentes, mener des analyses prédictives et proposer de nouveaux services, en particulier aux annonceurs professionnels. C'est le cas de l'Atelier business, une nouvelle offre de publicité personnalisée et géolocalisée pour les TPE-PME en France.



**Roland MINNE**  
Sales Director

+44 (0) 1932 895024  
info-uk@attunity.com  
3000 Hillwood Drive - Hillwood Business Park KT16  
ORS Chertsey UK

CONTACT



www.attunity.com



## BIG DATA MANAGEMENT AND DATA INTEGRATION SOLUTIONS

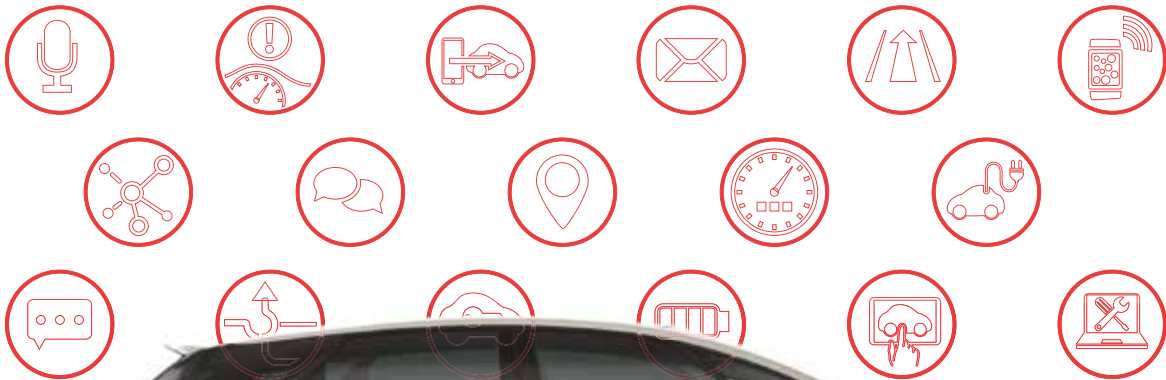
Provides Universal Big Data Availability to a Fortune 100 Automotive Maker

### Results

- Consolidating a global Data Lake
- Centrally monitoring all Replication tasks
- Minimising labour and reducing costs
- Realising faster insights and competitive advantage

### Additional Benefits

- Historical Insight
- Scheduled maintenance
- Design innovation
- Insurance premiums
- Pricing models
- Warranties
- Real-time Analytics
- Predictive maintenance
- Connected city
- Infotainment
- Autonomous driving
- Route optimisation



4,500 OLTP, ERP, CRM Systems

Right Data.  
Right Place.  
Right Time.

## Data Integration & Big Data Management

- Hadoop & Big Data
- Real-time streaming data into Kafka
- Databases & Data Warehouses
- On premise & in the cloud



**Thierry LALANDE**  
Senior Manager

+33 (0)1 58 86 30 00  
thierry.lalande@bearingpoint.com  
Immeuble Galilée - 51 esplanade du Général de Gaulle  
92907 Paris la Défense Cedex France



# BearingPoint®

## AMÉLIORATION DE LA LUTTE CONTRE LA FRAUDE GRÂCE AU BIG DATA UNE COLLABORATION DE LA CNAMTS ET DE L'ÉQUIPE DATA & ANALYTICS DE BEARINGPOINT

### À PROPOS DE LA CAISSE NATIONALE DE L'ASSURANCE MALADIE DES TRAVAILLEURS SALARIÉS (CNAMTS)

La Caisse nationale de l'assurance maladie des travailleurs salariés (CNAMTS) définit la politique de l'assurance maladie en France et pilote les organismes chargés de la mettre en œuvre. Elle assure notamment la gestion des facturations de l'ensemble des 90 familles de professionnels de santé (infirmiers, cardiologues, pharmacies, laboratoires...) et effectue des contrôles pour chacune d'elles, afin de vérifier la réalité des prestations facturées et, en cas de fraude, entamer des poursuites contre le professionnel de santé.

L'objectif principal de la CNAMTS est, au moyen de contrôles contraints, d'augmenter significativement la proportion de contrôles positifs – soit ceux qui permettent de détecter des comportements frauduleux.

Afin de répondre à cet objectif ambitieux, la CNAMTS a mis en place des méthodes statistiques de « datamining raisonné » depuis 2006, permettant d'optimiser la détection de fraude.

Bien que performantes sur les cas de fraude les plus aberrants, elles ne permettent toutefois pas de faire ressortir des profils de fraudeurs plus discrets. Or, la CNAMTS s'est rendu compte que les profils des fraudeurs tendaient à évoluer et à se complexifier au fil des contrôles, ce qui les rendrait plus difficilement identifiables par les méthodes de datamining raisonné.

C'est ainsi dans une démarche d'innovation qu'elle a décidé d'entreprendre un projet d'expérimentation avec HyperCube®, la solution d'analyse de données de BearingPoint, afin d'anticiper cette évolution en explorant les cas de fraude moins flagrants.

### DES RÉSULTATS MESURABLES ET CONCRETS

L'étude menée par la CNAMTS avec HyperCube® en 2014 a consisté à identifier les profils à risque, parmi la famille des infirmiers, sur base des contrôles effectués les années précédentes.

Les règles d'HyperCube® ont servi de base à des contrôles réalisés sur le terrain, parmi les infirmiers installés et exerçants en activité. En partant d'un taux de fraude très bas (0,7 % de fraudeurs constatés), il fallait concentrer le phénomène afin de gagner en sensibilité et ne pas lancer les contrôleurs sur des pistes inutiles.

Les profils identifiés par ces règles comprenaient une proportion de fraudeurs plus de cent fois plus importante que le taux moyen constaté. Suite à cette expérimentation « analytique », la CNAMTS a mené en 2014 et 2015 plusieurs dizaines de contrôles à partir de deux profils. Elle a ainsi constaté par rapport à la population contrôlée un **taux de fraude avérée de 70 %** et des redressements moyens approchant les **40 000 euros par fraudeur**.

Ces résultats finaux démontrent la stabilité dans le temps et la rentabilité des résultats obtenus avec des méthodes de « machine learning » telle qu'HyperCube®. Il a été constaté que les résultats obtenus avec cette méthode permettaient de cibler beaucoup plus précisément la population à risque qu'avec d'autres méthodes de « datamining supervisé » que la CNAMTS a par ailleurs également expérimentées.

Il est ainsi prouvé que l'utilisation de nouveaux algorithmes de data science (ou machine learning), pour identifier les cas de fraude, permet au contrôleur d'améliorer de façon conséquente ses ciblage et son efficacité sur le terrain.

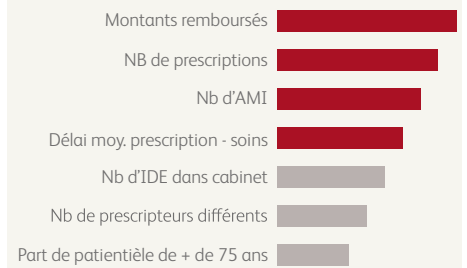
### LES FACTEURS CLÉS DE SUCCÈS EN 4 POINTS

- **La structuration d'une base de données** qui reflète le mieux possible le profil de chaque infirmier et les caractéristiques de son activité déclarée. À ce titre, les équipes de la CNAMTS, autour du docteur Pierre Fender, directeur de l'audit, du contrôle-contentieux et de la répression des fraudes, se sont attachées, pendant plusieurs semaines, à collecter et structurer les données pertinentes venues de nombreux systèmes d'information de la CNAMTS. Variables et facteurs influents.

#### Variables

-  152 indicateurs
-  16 183 infirmiers
-  15 variables apparues dans les règles

#### Facteurs influents



- **Le travail collaboratif** des experts fraude de la CNAMTS et des datascientists d'HyperCube®. Ce travail a permis, dans un premier temps, d'identifier et de prioriser les données à collecter et, dans un second temps, de sélectionner les profils opérationnellement pertinents parmi les profils mathématiquement fraudeurs.

- **Le caractère « explicatif » et « prédictif » de l'algorithme de machine learning HyperCube®.** Cet algorithme est capable de trouver des phénomènes locaux très petits et de les exprimer sous forme intelligible à des experts de la fraude qui ont pu s'attacher à comprendre les profils « mathématiques » et à les interpréter en comportements frauduleux. Cette étape de validation humaine du modèle est particulièrement importante pour la CNAMTS qui doit pouvoir justifier à tout moment les raisons d'une action publique. Ainsi les modèles prédictifs « boîte noire » comme ceux basés sur des « réseaux de neurones » ou des « forêts d'arbres décisionnels » n'apportent pas la transparence nécessaire.

- **La rigueur opérationnelle de la CNAMTS** qui a initié des tests importants sur le terrain. Elle a d'abord lancé en quelques mois plusieurs dizaines de contrôles à partir des profils identifiés, puis elle a attendu la réalisation de l'ensemble de ces contrôles, qui peuvent parfois durer plusieurs mois, enfin elle a fait le bilan des résultats de ces contrôles en termes de taux de fraudeurs constatés et de montants moyens de l'amende.

Pour de plus amples informations :  
<http://stories.bearingpoint.com/lutte-antifraude/>



# BearingPoint Data & Analytics

## Les offres clés de nos héros de la Data

### 1. Stratégie Big Data

- Visioning & Use Cases
- Business case & Partenariats

### 3. Analytics avancées

- Analytics : Descriptive, Prédicative & Prescriptive
- Data Visualisation & Dashboards

### 2. Data Governance & Architecture

- Schéma directeur BI/Data : technologie, architecture, solutions
- DataLab : Organisation, Processus, RH

### 4. Industrialisation & Transformation

- Data driven Apps
- Organisation & Transformation métier/IT

#### Nous contacter :

Thierry Lalande  
+33 6 20 84 54 92  
thierry.lalande@bearingpoint.com

Marc Hispa  
+33 6 21 01 02 06  
marc.hispa@bearingpoint.com

Immeuble Galilée  
51, esplanade du Général de Gaulle  
92907 Paris La Défense Cedex  
www.bearingpoint.com



**BearingPoint®**



**LA PLATEFORME DATA MANAGEMENT & EXCHANGE DE BLUE DME EST UNE SOLUTION METIER PERMETTANT AUX STATISTICIENS, ACTUAIRES, ANALYSTES RISQUES ET AUX DATA SCIENTISTS DE TROUVER FACILEMENT LES DONNEES LES PLUS UTILES ENFOUIES DANS VOTRE DATA LAKE ET DANS L'OPEN DATA.**

Notre plateforme **Data Management & Exchange (DME)** s'adresse aux **Chief Data Officers (CDO)** et aux **équipes métiers** pour leur permettre de révéler le potentiel des données internes dans les différentes entités d'une grande organisation, d'accéder à de nouveaux jeux de données externes pertinents automatiquement, et de développer des mécanismes de valorisation de données (écosystème de partenaires).

Après avoir accompagné de grands acteurs des secteurs banque, assurance, automobile et industrie dans la mise en place d'offres de services d'analyse prédictive, de machine learning et de modélisations prédictives (optimisation d'algorithmes de pricing, modèles de scoring, etc.), Blue DME a développé une solution innovante permettant de mettre les données du Data Lake et données internes directement dans les mains des équipes métiers.

Le DME, s'appuyant notamment sur des technologies Hadoop/Yarn, Spark et Elasticsearch, est composé des briques fonctionnelles suivantes :

- **Outil de gouvernance pour fédérer les data lakes** des entités d'une organisation ou entre organisations (garder une vision centrale fédérée des sources de données anonymisées et permettre de valoriser ses données en interne ou en externe) ;
- **Catalogue collaboratif orienté métier et moteur de recherche intelligent de données multi-sources** (données internes, de parcours client, open data, scrapées ou données externes payantes de partenaires) permettant un fort degré de capitalisation entre utilisateurs (retours utilisateurs, partages, listes thématiques personnalisables) ;
- **Outil de sélection unifiée de données pour sélectionner automatiquement les jeux de données pertinents dans le catalogue** (améliorer un score, rechercher des variables prédictives) et basculer facilement vers un outil de Data mining ou un studio de Data Science ;
- Solution permettant, aux application analytics métiers en production, d'**accéder aux données à jour du catalogue** (real time data flow).

Le numéro 1 de l'assurance-crédit dans le monde, Euler Hermes, a pu déterminer l'apport des données externes dans la prédiction du risque crédit pour accompagner les meilleures PME et startups européennes. L'équipe

datascience d'Euler Hermes a ainsi amélioré ses modèles en utilisant des données open data et des données externes non financières, puis a communiqué les résultats des analyses aux équipes de direction métier.

Un autre leader de l'assurance, composé de différentes marques / entités internes, a déployé la plateforme Data Management & Exchange auprès d'une centaine d'utilisateurs manipulant déjà des données : data scientists, chargés d'étude statistiques, experts des Systèmes d'Information Géographiques, experts risques, experts de la connaissance client, actuaires, etc. Cette entreprise souhaitait que tous ces utilisateurs, dont les acteurs métier, puissent utiliser au quotidien les données externes. La connaissance des données externes (notamment open data) ainsi que de beaucoup de jeux de données internes (exemples : les données de parcours client pour des acteurs autres que les experts connaissance client, ou les données géographiques travaillées par les experts SIG non utilisées par les autres collaborateurs du groupe) était bien sûr très diverse pour chaque profil utilisateur. L'objectif était donc bien :

- d'avoir accès à un catalogue de données intelligents avec une dimension collaborative forte pour **accompagner les utilisateurs n'ayant jamais utilisé des données externes**,
- et de laisser les profils ayant déjà une expérience utiliser les fonctionnalités les plus avancées (améliorer un score, réaliser un enrichissement) mais aussi **partager les retours d'expériences sur certains jeux de données**.

Les utilisateurs au sein de cette société d'assurance peuvent **automatiquement trouver les jeux de données pertinents pour leurs objectifs métiers** et manipuler aussi bien des données d'un data lake, que des données open data par exemple. Chaque groupe d'utilisateurs, proche de la notion d'entité métier, est réuni dans un espace de travail pour prendre en compte des contraintes réglementaires ou simplement échanger avec d'autres profils similaires et peut partager des nouveaux jeux de données au catalogue. La plateforme DME permet de **quantifier l'usage des jeux de données et révéler leur valeur : la première étape de la Data Monetization Interne**.

Le potentiel de valeur pour les organisations commence donc en interne (contribution des entités à la valeur économie de l'entreprise et partage des jeux de données et expériences) et peut se poursuivre en externe avec un **écosystème de partenaires qui souhaitent collaborer ensemble** sur des données qu'ils souhaitent partager de manière anonymisée.

Blue DME a implémenté ces services, principalement pour des entreprises banque / assurance et automobile, et **a amélioré significativement l'efficacité opérationnelle des équipes de data mining et de data science** (notamment pendant l'étape de data preparation qui occupe plus de 60% de leur temps) et **permet aux CDO et DSI de valoriser les investissements réalisés dans leurs data lakes**.



Service intelligent de recherche de données externes améliorant un score existant



Service intelligent de recherche de variables prédictives



# Blue DME

WE BUILD THE DATA DRIVEN BUSINESS

## # Data Management & Exchange

Amener les données de votre Data Lake  
au cœur du business



### RECHERCHE INTELLIGENTE DE DONNEES

Révéler les jeux de données pertinents internes, découvrir ceux provenant de l'open data et de vos partenaires, et ainsi améliorer des modèles prédictifs tout en gagnant en efficacité opérationnelle.



### GOVERNANCE DES DONNEES

Fédérer les data lakes de vos entités et filiales pour développer une stratégie Data globale respectant les contraintes réglementaires et vos besoins de flexibilité.



### COLLABORATION DES EQUIPES METIERS SUR LA DATA

Développer la capitalisation autour de la connaissance des jeux de données pour améliorer les recommandations.



### DATA MONETIZATION INTERNE

Identifier la contribution métier des données de vos entités et filiales à votre entreprise, et les partager avec vos partenaires.

# Banque

# Assurance

# Automobile

# Industrie

# BLUESCALE

**Franck BELLENGER**

Directeur associé

+33(0)1 53 25 02 10

contact@bluescale.com

55 rue du Faubourg Montmartre 75009 Paris France



Analytique



API



Applications



Consulting/CSII  
Intégrateurs



Dataviz



Infrastructure



Institutionnels



[www.bluescale.com](http://www.bluescale.com)



bluescale

## BLUESCALE AIDE SES CLIENTS À METTRE EN ŒUVRE LES NOUVEAUX LEVIERS D'INNOVATION TOUT AU LONG DE LA CHAÎNE DE VALEUR DU BIG DATA

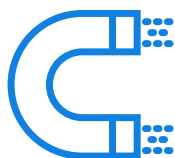
Bluescale est composé d'une équipe unique de 50 consultants, développeurs et ingénieurs de la donnée, maîtrisant les dernières innovations et les nouveaux paradigmes de la data, avec pour seul objectif de mettre en place de façon industrielle et agile les briques technologiques de la chaîne d'information de demain.

### Bluescale intervient tout au long du cycle de vie de vos projets :

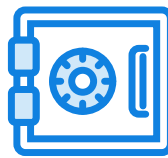
- Accompagnement fonctionnel et gouvernance de la donnée
- Conception et mise en œuvre d'architectures big data assurant performance, qualité, sécurité, traçabilité, agilité et innovation
- Mise en place de processus d'industrialisation de prototypes ou de modèles
- Intégration de solutions de dataviz, de mise en valeur et partage de la connaissance
- Développement d'applications agiles Data Driven opérationnelles
- Conduite du changement, formations

**Bluescale, c'est plus de 10 ans d'expérience dédiés à la data et à l'innovation.**

## WE ARE DATA ENGINEERS



Collecte et  
Processing



Organisation  
des données



Architecture  
Data



Valorisation  
de la donnée

# Plateforme temps réel de suivi des clients en magasin

## ÉTUDE DE CAS



**SES-IMAGOTAG EST LE LEADER MONDIAL DES ÉTIQUETTES ÉLECTRONIQUES. DANS UN CONTEXTE DE TRANSFORMATION NUMÉRIQUE DU COMMERCE PHYSIQUE, SES-IMAGOTAG DOIT AUJOURD'HUI PROPOSER AUX ACTEURS DU MONDE DE LA DISTRIBUTION DES SOLUTIONS DIGITALES INNOVANTES.**

Bluescale a mis en place depuis début 2015 plusieurs solutions technologiques permettant d'atteindre un double objectif. Le premier objectif est d'enrichir l'expérience client en magasin, en offrant le meilleur du retail et du digital : consultation des fiches produits avec le smartphone via l'étiquette NFC, analyse des consultations et retargeting, alertes sur les produits (évolution du prix, nouveaux commentaires, etc.), digitalisation des files d'attente, rapprochement des données de stocks en ligne et en magasin. Le second objectif est de devenir le « Google Analytics © » des points de vente en offrant aux retailers des compteurs de présence en temps réel, une fiche client enrichie (produits consultés, historique de visites), des tableaux de bord statistiques sur le parcours client, leur fidélité, les taux de conversion, une détection de présence (pour préparer un retour SAV dès l'entrée du client en magasin par exemple).

### ARCHITECTURE MISE EN PLACE

Deux grands enjeux se sont immédiatement posés en terme d'architecture :

1. La plupart des services nécessitent une mise à disposition immédiate des données.
2. Afin de répondre aux besoins spécifiques des enseignes mettant en place la solution, il est nécessaire de pouvoir déployer rapidement de nouveaux services personnalisés.

D'un point de vue technique, la plateforme s'appuie sur deux axes forts : tout d'abord l'ingestion en temps réel des données, quelle que soit la source, associée à un collecteur permettant d'historiser ces informations, ce qui permet la mise en place d'applications d'analyse en temps réel. Ensuite, la mise en place de microservices, offrant une grande modularité dans le déploiement de nouvelles fonctionnalités.

### QU'EST-CE QU'UNE ARCHITECTURE LAMBDA ET POURQUOI AVOIR CHOISI CETTE SOLUTION ?

Bluescale a préconisé et mis en place une architecture dite « lambda » composée de trois couches :

- une couche temps réel : toutes les données produites sont distribuées dans des flux. Des applications d'analyse peuvent les consommer selon leurs besoins et les cas d'usages (compteurs, détection de présence, alertes).
- une couche batch : les données collectées en temps réel sont historisées et accessibles. Des traitements d'analyse plus coûteux et donc moins fréquents génèrent des données raffinées, qui sont ensuite exposées à la couche de service.
- une couche service : celle-ci expose les données enrichies aux clients.

Cette architecture permet une grande agilité dans la mise en place de nouveaux services ou l'intégration de nouvelles sources de données.

Cette richesse et cette flexibilité impliquent une vigilance accrue sur la gouvernance des données, compte tenu de la multiplication des flux et des services. Pour répondre à cette attente, l'architecture s'appuie entre autre sur le pattern « Service Registry », qui offre un référentiel et une gouvernance des cycles de vie des services pour les applications.

Lorsque SES-imagotag a décidé d'ajouter la technologie 2,4 GHz à son offre de communication avec les étiquettes électroniques, le Wi-Fi est apparu comme un incontournable demandé par les clients.

Après plusieurs mois de coopération avec un fabricant allemand, la société a commercialisé le premier point d'accès mixte : pilotage d'étiquettes électroniques + Wi-Fi avec en prime le système LBS (Location Based Services). Ce système « écoute » les demandes de connexion des smartphones de passage et enregistre de façon anonyme ces informations à un rythme incroyable : plusieurs millions par jour.

Confrontés à cette problématique de big data sur deux V (Volume, Velocity mais pas Variety), nous avons fait appel à la société Bluescale pour nous aider à identifier la stack de technologies capable d'encaisser la charge en temps réel, de l'analyser en volume (plusieurs centaines de magasins) et de la restituer de façon moderne c'est-à-dire en responsive design.

Aujourd'hui non seulement l'outil fonctionne selon le cahier des charges, en plusieurs langues, mais en plus il a permis des évolutions spécifiques au métier, en particulier l'ouverture de l'outil aux annotations, une fonctionnalité souvent oubliée par les outils de business intelligence classiques.

Enfin il apparaît maintenant que le HTML5 ne suffit plus pour une navigation aussi fluide que dans une app native et nous envisageons de porter l'ensemble dans un framework capable d'ajouter cette touche UI/UX qui fait la différence auprès des utilisateurs. Naturellement nous comptons sur Bluescale pour nous accompagner pour cette nouvelle étape.



Jean-Christophe SOLUS / SES-imagotag

**Gilles RICHARD**  
Directeur Associé

+33(0)1 53 48 15 57  
grichard@bluesoft-group.com  
14 place de la Coupole - 94220 Charenton Le Pont



[www.bluesoft-group.com](http://www.bluesoft-group.com)



## BLUESOFT GROUP, ESN DE 400 CONSULTANTS, ACCOMPAGNE SES CLIENTS DANS LA TRANSFORMATION DIGITALE DE LEURS MÉTIERS.

Présents en régions et à l'international, précurseurs dans la mise en œuvre de solutions techniques et fonctionnelles pour les usages de demain, nous investissons fortement en innovation, R&D, notamment via notre laboratoire BlueDsX (Data Science - Data - Résilience des SI), et en formations certifiantes à destination de nos salariés.

## TROIS ACTIVITES MAJEURES

- Data Management (40%)
- Accompagnement à la transformation digitale (30%)
- Résilience et sécurité des S.I (30%)

### Le Design Thinking au service du Big Data

Avec des spécifications de cas d'utilisation-métier pertinents et opportuns

### Des Proof of Concept ciblés et précis

Avec un engagement de vos ressources, progressif, itératif et raisonné

Notre offre *Blue Design*, c'est

### Big Maturity Model et Audit

Avec notre modèle Big Maturity Model (BMM), nous évaluons la maturité de vos projets Big Data et déterminons les opportunités d'améliorations

### Des infrastructure résilientes

Avec notre expertise reconnue en performance des infrastructures et en création d'algorithmes dimensionnés au plus près des exigences



## PRESTATIONS

Leader de la DATA, au cœur des SI, nous travaillons depuis 10 ans dans l'adaptation des infrastructures et des middlewares aux nouveaux besoins fonctionnels. Nous définissons et mettons en œuvre des solutions performantes, réduisons les coûts d'exploitation et de maintenance. Nos experts opérationnels ou stratèges assistent leurs clients sur des sujets critiques. Nous maîtrisons la transformation digitale et ses outils (cloud, Big Data, IoT ...)

## NOTRE ACTUALITE

De l'analyse massive de logs sur une plate-forme Big Data dédiée (systèmes, réseaux, stockage, applications, etc.), via des techniques issues de l'analyse de langage (NLP), de la recherche de patterns (clustering) et machine learning (entropie partagée et arbres de décision), nous détectons pour nos clients les prémisses de compromissions ou d'incident réseaux et legacy.

# *Blue Design*

L'ALIGNEMENT STRATEGIQUE ET  
OPERATIONNEL DE VOS PROJETS  
BIG DATA

**Yann MANEIX**

Directeur Technique

+33(0) 6 31 70 74 65

partner@bmb-services.com

11 rue Bargue 75015 Paris France



Que ce soit au travers d'un smartphone, des réseaux sociaux, ou de votre voiture, notre environnement, de plus en plus digitalisé, stocke d'énormes quantités d'informations, structurées ou non. Les entreprises, au travers de cette transformation numérique, cherchent à exploiter au mieux ce capital de données afin de le transformer en avantage concurrentiel. Réinventer son offre et améliorer le time-to-market devient essentiel pour se démarquer sur son marché.

BMB services est un cabinet de conseil qui accompagne ses clients dans la transformation du système d'information autour des innovations SAP. Des entreprises utilisatrices nous sollicitent aujourd'hui dans l'élaboration de cas d'usage, ou dans la mise en œuvre de prototypes sur des scénarios déjà conceptualisés.

A ce jour, la mise en œuvre de plateformes Big data en est encore à ses débuts et se réduit souvent par l'installation de systèmes d'archivage (data lake). La construction de modèles d'analyse et de corrélation des données est l'étape suivante de cette transformation. Les possibilités amenées par les dernières avancées technologiques et les innovations permettent de réagir en temps réel. Construire une stratégie prédictive pour ses clients est au cœur de nos préoccupations : on ne se tourne plus uniquement vers le passé mais on agit en temps réel pour influencer sur l'avenir.

La connexion entre les systèmes SAP de l'Entreprise et les plateformes Hadoop est en passe de devenir une des préoccupations majeures de notre écosystème clients. L'éditeur SAP s'implique dans cette nouvelle orientation et s'associe à des acteurs majeurs du Big Data pour proposer des solutions intégrées à ses clients. Grâce aux produits comme VORA et Smart Data Access, il devient natif de mettre en relation la base de données in-memory SAP HANA et Hadoop pour pouvoir, au sein du SI, mettre en corrélation des données hétérogènes (structurées et non structurées).

Les entreprises sont, par conséquent, de plus en plus demandeuses de Proof-of-Concept ou prototypes contextualisés sur les possibilités d'intégration de données externes, souvent non structurées dans un système SAP existant. Nos connaissances des industries, des processus métiers et de nos clients nous permettent de proposer une intégration et une modélisation de données au plus proche de leurs besoins métiers, en s'appuyant sur des outils puissants et innovants. Cela touche à la fois les données fournies à un instant T mais également l'analyse prédictive.

### EXEMPLES D'USAGES EN ANALYSE PRÉDICTIVE :

Dans la collecte des données issues des tickets de caisse, une entreprise de concessions de restaurants peut déterminer de façon précise quels sont les produits les plus consommés, les matières premières les plus utilisées par zone géographique. La corrélation avec des données pu-

bliques telles que les prévisions de trafic permet d'optimiser la gestion de ses stocks et minimiser les coûts d'approvisionnement.

Un groupe spécialisé dans la location de vélo libre-service peut, grâce à l'analyse des données non structurées issues des réseaux sociaux, et à l'étude des statistiques d'utilisation de ses bornes, prévoir l'implantation d'une nouvelle station vélo. La connaissance des taux d'utilisation par zone géographique combinée à l'étude des commentaires sur les réseaux sociaux permet de segmenter la demande et de mieux répondre aux exigences des clients consommateurs finaux.

Outre ces scénarios qui permettent d'anticiper et construire une stratégie d'entreprise basée sur des prévisions, le Big Data permet aussi des usages opérationnels de pilotage en temps réel.

### EXEMPLES D'USAGES EN TEMPS RÉEL :

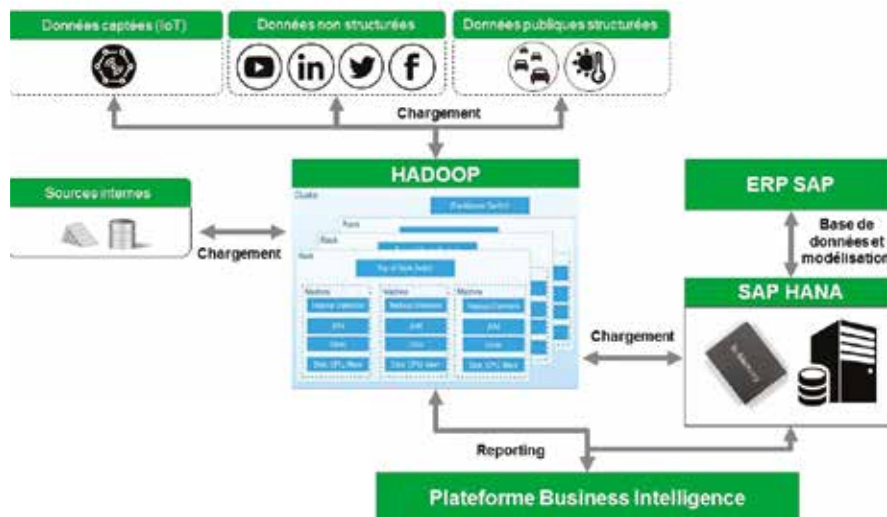
La mise en place de capteurs sur les points de vente des gares, aéroports ou stations-service pour gérer l'approvisionnement de produits alimentaires dans les rayons permet au travers d'une remontée d'alerte en temps réel, d'être informé et de pouvoir déclencher le réapprovisionnement. Cette technique permet d'éviter des ruptures de stock dans les rayons et tendre ainsi vers l'excellence opérationnelle.

Une application industrielle réalisée par une entreprise dans l'aéronautique consiste à optimiser la maintenance des machines au moyen de capteurs, réglés pour déclencher une intervention et ainsi éviter une panne ou un remplacement selon l'usure des pièces détachées détectée par le capteur.

Pour une entreprise leader dans la publicité urbaine, la capacité à capter les informations des smartphones des passants à proximité d'un panneau permet de cibler, selon les habitudes de consommation de chaque personne, les offres promotionnelles susceptibles de l'intéresser par rapport à ses besoins. La relation commerciale devient individuelle et personnalisée.

Sur un marché où l'explosion de la data complexifie l'interprétation et l'exploitation des messages, ce sont autant d'usages qui permettent pour une entreprise de s'améliorer et de se renouveler constamment.

En veille technologique permanente, BMB Services, cabinet de conseil leader de la transformation sur SAP, continue de renforcer sa maîtrise des processus industriels autour de l'énergie, la restauration collective, l'aéronautique, ou le monde médical, et met aujourd'hui son savoir technologique au service de ces industries en proposant des modèles d'exploration, de data-mining et de prédiction de l'information lors d'événements ou de démonstrations portés par de l'IoT (capteurs).







## NOTRE APPROCHE

**Aligner technologies et métiers  
pour créer plus de valeur**

# PRÉVOYEZ AVEC LE BIG DATA

*Offrez-vous de la haute performance  
et de l'agilité sur vos données avec  
les technologies Hadoop® du moment.*

Faîtes-vous accompagner  
dans votre orientation stratégique  
pour mieux analyser, prédire  
et comprendre l'information qualitative  
de demain.

## CONTACTEZ-NOUS

Stéphane Parisis

**Directeur associé**

Mobile : 06 22 61 87 64

[stephane.parisis@bmb-services.com](mailto:stephane.parisis@bmb-services.com)

William Grosjean

**Directeur associé**

Mobile : 06 84 77 02 78

[william.grosjean@bmb-services.com](mailto:william.grosjean@bmb-services.com)



Bâtiment B  
11, rue Bargue  
75015 Paris- France



[www.bmb-services.com](http://www.bmb-services.com)

## BUSINESS & DECISION

**Mick LEVY**

Directeur Business Innovation

+33(0)1 56 21 21 21

bigdata@businessdecision.com

153 rue de Courcelles 75017 Paris France



Analytique



API



Applications



Consultants/ESM/  
Intégrateurs



Dataviz



Infrastructure



Institutionnels



www.businessdecision.fr

**Business & Decision**

### BUSINESS & DECISION : DATA EMPOWERED

Business & Decision est un acteur historique et de référence dans le large domaine de la Data.

Présent dans 15 pays, Business & Decision emploie plus de 2 500 personnes en France et dans le monde. Fort d'une expertise unique sur ses compétences historiques (Business Intelligence, CRM, e-Business), le groupe s'appuie sur un important réseau de partenariat avec les éditeurs logiciels les plus innovants du marché.

Avec l'innovation comme crédo, Business & Decision s'est imposé, en 20 ans, comme un acteur français de référence majeur dans la Data et le Digital.

#### L'activité de Business & Decision se décline sur deux grands domaines d'activité :

- Data : Big Data, Advanced Analytics, Data Science, Webanalytics, Dataviz, Data Management Platform (DMP), Business Intelligence, Référentiels MDM et Qualité de Données, Cloud...

- Digital : Parcours clients, Relation et connaissance clients, Digital Marketing, Transformation Digitale, e-business, Mobilité et objets connectés, Social, Communication Digitale, Cloud...

Dans ses domaines de spécialisation, Business & Decision accompagne ses clients sur toutes les étapes des projets au travers de prestations de conseil, de choix de solution, de conception et développement d'applications, de formation et d'accompagnement au changement, d'hébergement et plus largement, de gestion complète du capital applicatif.

### ACCELERATEURS PROJETS

**Business & Decision propose des approches packagées pour accélérer la mise en œuvre des solutions Big Data :**

- Data IoT pour ne pas rater l'enjeu-clé de votre initiative IoT : le traitement de la Data
- Jumbo#1 pour réaliser en quelques jours une initiative Big Data
- DataScience4Business pour réaliser une étude ou un projet statistique ou prédictif
- LOD.Eolas (Open Data) pour ouvrir les données de votre organisation et participer à l'open-innovation

Le groupe offre également des services dans le domaine du Cloud Computing en proposant des solutions d'hébergement en France de type Big Data as a Service.

Fondées sur l'expérience et le savoir-faire, les expertises de Business & Decision sont régulièrement publiées, notamment dans le blog d'experts Big Data et Digital (<http://blog.businessdecision.com>) ainsi que lors de conférences et événements.

*ROI<sup>2</sup> : Return On Investment through Return In Information*

FAITES APPEL AUX  
**DATA-HEROES**  
POUR VOTRE **BIG DATA**



**Yamina KHELIFI**

Chargée d'affaires SI

+33(0) 1 41 13 10 95

yamina.khelifi@centralesupelec.fr

Grande voie des Vignes 92295 CHATENAY MALABRY  
France



[www.exed.centralesupelec.fr](http://www.exed.centralesupelec.fr)



CentraleSupélec



CentraleSupélec Exed est l'entité dédiée à la formation continue de CentraleSupélec. En accompagnant les professionnels aux différentes étapes de leur carrière dans le champ des sciences de l'ingénieur et du management, nous prolongeons notre mission première : celle de former les leaders de culture scientifique et technologique, et des innovateurs, capables de relever les grands défis de leur temps.

Les formations qualifiantes et diplômantes proposées par CentraleSupélec Exed s'articulent autour de quatre grands types de formations : les Formations Mastère Spécialisé®, les Executive Certificates, les formations courtes et les formations sur-mesure.

### LA FORMATION BIG DATA – ENJEUX ET OPPORTUNITÉS

À l'ère du numérique, le déluge de données crée de nouvelles opportunités économiques pour les entreprises. La possibilité d'analyser ces masses de données représente un avantage non négligeable en termes de compétitivité, et de nouveaux métiers apparaissent comme celui de Data Scientist. Toute la chaîne du décisionnel est impactée par ces « big data », de la collecte des données hétérogènes à leur analyse et leur visualisation, en temps contraint. Différents métiers de l'entreprise sont également impactés par ce phénomène comme le marketing ou la stratégie d'entreprise. L'objectif de cette formation est de dresser un état des lieux et un panorama des technologies disponibles pour permettre aux entreprises de garder la maîtrise de ces données et de la valeur qu'elles peuvent générer, et ainsi, de gérer la transformation vers l'entreprise numérique. L'organisation centrée données permet ainsi de placer les données au centre de la gestion opérationnelle et décisionnelle, et de réduire le temps de prise de décision.

### OBJECTIFS DE LA FORMATION

- COMPRENDRE les enjeux des Big Data
- IDENTIFIER les opportunités à travers l'open data et les cas d'usage métiers
- ACQUÉRIR le vocabulaire lié à cet écosystème complexe
- ADAPTER la stratégie d'entreprise
- MAÎTRISER les enjeux juridiques et liés à la protection des données
- ACQUÉRIR les technologies des Big Data (acquisition et stockage des données, traitement distribué, analyse de données à large échelle)
- MAÎTRISER les technologies par des études de cas concrètes
- CONCEVOIR un projet Big Data

### EN SAVOIR PLUS...



## ÉLEVER SES COMPÉTENCES POUR RELEVER LES NOUVEAUX DÉFIS D'AUJOURD'HUI !



### EXECUTIVE CERTIFICATE **BIG DATA** POUR L'ENTREPRISE NUMÉRIQUE



Enregistré au titre  
RNCP Niveau I  
Expert en ingénierie  
numérique  
Code CPF 178043

Enjeux, perspectives et panorama des Big Data / Aspects juridiques et protection des données  
Acquisition et stockage des données / Traitement distribué / Analyse des données et Big Data,  
confess your Data / Cadrage et management de projets Big Data

+ DES FORMATIONS SUR MESURE, PERSONNALISÉES, ADAPTÉES À VOS BESOINS !



[www.exed.centralesupelec.fr](http://www.exed.centralesupelec.fr)  
[info.exed@centralesupelec.fr](mailto:info.exed@centralesupelec.fr)  
+33 (0) 141 131 500

Le monde change  
Nous vous aidons à changer



SYSTÈMES  
D'INFORMATION  
ET DIGITAL

EXPERIENCE



**DAMIEN DROISY ET RAPHAEL BECHU. LE PREMIER TRAVAILLE DANS LE SECTEUR BANCAIRE, LE SECOND DANS CELUI DES TRANSPORTS. TOUS DEUX ONT SUIVI LA FORMATION CERTIFIANTE « BIG DATA POUR L'ENTREPRISE NUMÉRIQUE » DE CENTRALESUPÉLEC EXED. À CHACUN D'ENTRE EUX, NOUS AVONS POSÉ TROIS QUESTIONS.**

**POUR QUELLE RAISON AVEZ-VOUS CHOISI DE SUIVRE CETTE FORMATION ?**

**Damien DROISSY :** Pour moi comme pour mon employeur, comprendre et mettre à profit le Big Data n'était pas une option, mais une obligation. Une nouvelle équipe se mettait en place avec pour mission de propulser l'entreprise au cœur même de l'ère numérique en tirant le meilleur parti du Big Data. Rejoindre cette équipe était un objectif personnel. Par intérêt, bien sûr, mais aussi par souci d'évoluer de façon constante dans mon cœur de métier.

**Raphaël BECHU :** Chef de projet, j'ai rapidement pris conscience pour mon entreprise des opportunités à saisir. Le sujet étant technique et spécifique il fallait trouver une formation qui me permettrait de comprendre les enjeux d'un projet Big Data mais aussi d'en maîtriser les aspects et ses approches. Nous en avons discuté avec mon employeur et la formation de CentraleSupélec Exed s'est imposée.

**QU'EST-CE QUE CETTE FORMATION A CHANGÉ DANS VOTRE VIE PROFESSIONNELLE ?**

**Damien DROISSY :** Je participe aujourd'hui activement à l'implantation du projet Big Data dans mon entreprise. J'ai intégré l'équipe Big Data et ce changement d'affectation me place aux premières loges de son intégration progressive dans les activités de l'entreprise.

**Raphaël BECHU :** La formation Certifiante de CentraleSupélec Exed m'a doté de nouveaux outils et de nouvelles compétences, que j'utilise au quotidien. Le Big Data est en phase de développement et d'intégration et je ne suis pas peu fier de participer à cette évolution en tant qu'acteur et non spectateur.

**OÙ EN ÊTES-VOUS AUJOURD'HUI DANS VOTRE PARCOURS ?**

**Damien DROISSY :** Suivre cette formation a été un tremplin pour aller plus loin encore dans la création de valeur pour mon entreprise...et pour moi-même. J'ai vraiment pris conscience de l'importance d'utiliser ma banque d'heures de formation pour maintenir mon employabilité

**Raphaël BECHU :** Aujourd'hui je mets à profit mes compétences transversales pour travailler au développement d'une stratégie Big Data intégrée dans mon entreprise. Je tiens compte de nombreux enjeux liés au traitement et à l'utilisation des données. Les implications juridiques, la structure et le fonctionnement du réseau de distribution, la stratégie de marketing de produit ou de service de l'entreprise, sont autant de paramètres à prendre en compte et que nous avons abordés dans la formation de CentraleSupélec Exed.

Le Big Data étant au cœur du développement des entreprises, tous secteurs confondus, la formation dispensée par CentraleSupélec Exed est stratégique. Pour les entreprises, bien sûr, mais aussi et surtout pour ceux qui sont conscients que c'est maintenant que les besoins se font sentir et qu'il y aura forcément ceux qui seront « in » et ceux qui seront « out ».

La réponse de CentraleSupélec Exed s'est rapidement mise en place, l'Executive Certificate Big Data pour l'entreprise numérique permettant aux participants d'acquérir rapidement tous les savoirs et toutes les compétences requis pour qu'au lendemain même de leur formation ils puissent être plongés dans l'action et travailler concrètement aux projets d'avenir.





## SIMPLIFIER L'ACCÈS AUX DONNÉES POUR UN DÉVELOPPEMENT PLUS RAPIDE DE NOUVEAUX MÉDICAMENTS



## PFIZER RÉDUIT DE MOITIÉ LE TEMPS DE DÉVELOPPEMENT DE SES PROJETS AVEC LA SUITE CISCO DATA VIRTUALIZATION

Au sein du département informatique de Pfizer, les chercheurs et responsables de portefeuilles produits collectent et analysent un ensemble complexe de données en provenance de sources multiples et disparates : des bases de données scientifiques, intégrant des diagrammes de structures moléculaires, des résultats de tests cliniques antérieurs, l'état d'avancement du projet, le nombre de chercheurs impliqués, le coût du personnel.

L'intégration de ces données hétérogènes représentait pour Pfizer un goulot d'étranglement empêchant de fournir de l'information efficace à ses équipes durant le processus de découverte de nouveaux traitements.

Pour gérer l'intégration de ses données, Pfizer s'est initialement basé sur plusieurs approches : le développement de codes personnalisés entre les sources de données et les applications les plus consommatrices, le processus classique d'extraction, de transformation et de chargement (ETL) pour créer des entrepôts de données et des data marts.

La complexité croissante du Big Data réduisait sensiblement l'efficacité du codage manuel. En outre, la réplique de données avait un impact significatif sur la charge de travail des équipes informatiques redoublant d'efforts pour maintenir la qualité et la sécurité des données répliquées. De plus, si l'ETL permettait l'intégration des données de manière efficace pour mettre en œuvre des analyses multidimensionnelles et à large échelle, son développement requérait des mois, avec, au final, des données souvent obsolètes.

Enfin, aucune de ces approches traditionnelles ne s'intégrait de manière fluide dans la nouvelle stratégie SOA de Pfizer mettant l'accent sur la création d'objets de données réutilisables.

Jonglant entre la complexité administrative et la difficulté à accéder aux sources sous-jacentes, Pfizer passait ainsi quatre mois à finaliser ses différentes phases de recherche & développement, de tests et de mise sur le marché de nouveaux médicaments.

### La solution

Cisco a implémenté un serveur d'information (CIS). Pilier de la Suite Cisco Data Virtualization, CIS permet d'accéder aux données de façon non invasive, fédère des données disparates et simplifie les données complexes. Il permet à Pfizer d'évoluer vers de la visualisation, de l'abstraction et de la fédération de données, à la fois en environnements SOA et non SOA et supporte dorénavant les exigences de développement et d'exploitation des plans de recherche scientifiques.

Au sein d'un environnement de développement intuitif, les données sont extraites sous forme de vues relationnelles pour la création de rapports ou encore des services Web SOA. Par ailleurs, le moteur de recherche haute performance facilite l'accès et fédère puis délivre les données distribuées aux applications consommatrices en temps réel. Les utilisateurs bénéficient ainsi de données en libre-service pour un processus de décision accéléré.

### Nouvelles capacités et bénéfices pour les équipes de Pfizer

- Un développement automatisé de données libérant du temps pour les développeurs tout en réduisant de moitié la durée totale de développement d'un projet,
- Un environnement de développement intuitif, de la sécurité intégrée et la génération automatique de web services, requérant moins de compétences spécialisées,
- Des services de données WSDL (Web Services Description Language) compatibles SOA fournissant des données sous la forme requise par les développeurs du portail web,
- Des services de données combinés, plus faciles à maintenir que des scripts ETL et de distribution lors de modifications des sources sous-jacentes ou du portail web,
- Des actifs de services de données sous forme d'objets entièrement réutilisables.

### Résultats

- Une réduction de 50% du temps de développement des projets, passant de 3-4 mois à seulement 6-8 semaines
- Une qualité des données améliorée de 5% via l'utilisation de services de données
- Des coûts d'infrastructure opérationnels réduits de 10%
- Une baisse drastique de l'ordre de 60% des risques de non atteinte des objectifs de mise sur le marché pour les nouveaux traitements

### Pour de plus amples informations

- [www.cs.co/data-virtualization](http://www.cs.co/data-virtualization)

### Comment le Big Data bénéficie à votre productivité

Selon IDC, le bénéfice tiré par les entreprises exploitant le Big Data sur des serveurs Cisco UCS pourrait s'élever à plus de douze millions de dollars sur trois ans. Cela passe par une productivité accrue des data scientists et des analystes, ainsi que par la vente de nouvelles offres ou services. Selon les entreprises consultées, les serveurs Cisco UCS contribuent à cette nouvelle valeur en accélérant le déploiement, en garantissant l'évolutivité, et en assurant les niveaux de performances de leurs solutions Big Data. Etude disponible en ligne via le QR code.



Vos données sont partout.  
Augmentez rapidement leur valeur métier.



ROI  
à 3 ans

**366%**

Rapidité d'accès  
au marché



**56%**

Gain de productivité Data  
Scientists / Analytique



**26%**

Source IDC 2016

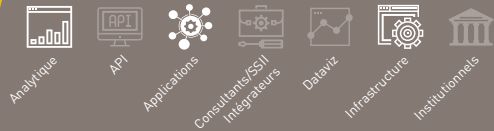


Retrouver les résultats  
de l'étude IDC :



**Romain PICARD**  
Regional Director SEMEA

+33(0)1 40 07 80 46  
info.semea@cloudera.com  
72 rue du Faubourg Saint Honoré 75008 Paris France



fr.cloudera.com



**cloudera**<sup>®</sup>  
Ask Bigger Questions

**LAISSEZ VOS DONNÉES CONTER UNE AUTRE HISTOIRE**

Imaginez tout ce que votre entreprise pourrait faire, si toutes vos données étaient rassemblées sur une même plate-forme centralisée, sécurisée et entièrement gérée, accessible n'importe quand et n'importe où par l'ensemble de vos services. Cloudera rend cela possible grâce à sa plate-forme moderne destinée à la gestion et à l'analyse des données.

Les données vous aident à résoudre les problèmes les plus complexes. Un Enterprise Data Hub (EDH) permet d'obtenir une transformation opérationnelle sur toute votre chaîne de valeur, de la conformité et la réduction des coûts à la connaissance des clients, en passant par les avantages compétitifs.

Cloudera propose la plate-forme basée sur Hadoop la plus rapide, la plus simple et la plus sécurisée au monde. Nous vous aidons à résoudre vos plus grands défis métiers.

**LES DONNÉES POUR TRANSFORMER VOTRE MÉTIER**



Vision du Client 360°



Améliorer l'efficacité des produits et services



Réduction des risques pour l'entreprise

**UNE PROPOSITION DE VALEUR UNIQUE DANS LE CLOUD**

**Élasticité**

La plate-forme de Cloudera peut tirer parti de l'élasticité du Cloud pour une évolution dynamique de vos clusters en fonction de vos pics d'activité.

**Multi-cloud**

Préserver la flexibilité de votre entreprise et minimiser le lock-in. Exécutez la même application ou les même Workload dans les principaux Cloud public du marché public ou dans votre propre cloud privé

**Un Cloud pour vos applications critiques d'entreprise**

Réduisez vos risques avec la plate-forme de gestion de données la plus agile, la plus sécurisée et assurant une gouvernance complète de vos données pour vos applications critiques en toutes conformité.





**DEVENEZ UNE ENTREPRISE PILOTEE  
PAR LES DONNEES AVEC CLUDERA**

**EVOLUTIVITE**

Une architecture évolutive vous permet d'accroître la capacité et les performances de manière linéaire en ajoutant simplement des serveurs

**FLEXIBILITE DU STOCKAGE**

Stockez tout type et volume de données dans son format original, sans modélisation de données

**FLEXIBILITE DES TRAITEMENTS**

Divers outils pour tirer toute la valeur des données – traitement batch, SQL analytique, recherche, machine learning et calculs statistiques

**FLEXIBILITE DU DEPLOIEMENT**

Déployez votre cluster sur des équipements hardware physiques, des clouds privés ou publics

**MAÎTRISE DES COÛTS**

Jusqu'à 90% plus économique que des solutions de gestion de données traditionnelles

**ARCHITECTURE OUVERTE**

Adoptez des standards ouverts et des composants open source pour bénéficier d'une qualité et d'une valeur durables, d'une compatibilité avec un écosystème étendu et de la portabilité pour réduire le risque de dépendance vis à vis d'un fournisseur

**INTEGRATION**

Capitalisez sur vos investissements existants pour accélérer l'adoption et réduire le coût total de possession

## CLUDERA ENTERPRISE :

### Hadoop pour l'entreprise

Cloudera Enterprise aide votre entreprise à démocratiser l'analytique à travers l'ensemble de l'organisation, avec Apache Hadoop, en tirant parti des meilleures technologies ouvertes dotées des fonctionnalités nécessaires pour répondre aux besoins métiers les plus sophistiqués. Conçu spécifiquement pour des environnements critiques, Cloudera Enterprise inclut CDH, la plate-forme open source basée sur Hadoop la plus populaire au monde, ainsi que les outils d'administration système les plus innovants.

La solution offre également l'accès à un support technique dédié et à la communauté de nos développeurs et experts Hadoop. Disposant de la flexibilité et de la stabilité nécessaires pour évoluer au même rythme que vos activités, Cloudera est votre partenaire pour tous vos projets de Big Data.

### Repenser la Gestion des Données

Cloudera Enterprise (reposant sur Hadoop) est une solution unifiée qui vous aide à stocker et à analyser l'ensemble de vos données et métadonnées, avec des fonctionnalités avancées de sécurité et de gouvernance conforme aux exigences réglementaires, et d'administration de systèmes de bout en bout. Grâce à Cloudera Enterprise, les entreprises sont en mesure de tirer le meilleur parti de leurs données pour bénéficier de capacités analytiques à l'échelle de toute l'organisation, optimiser leur stratégie et leur revenu, et réduire les coûts – tout en gérant efficacement les risques et en répondant aux contraintes réglementaires.

Cloudera Enterprise, dont le cœur repose sur Apache Hadoop, est :

**Unifié** — un système unique et intégré, offrant aux utilisateurs et aux applications l'accès à un entrepôt unique de données sur une infrastructure commune ; le déplacement des données n'est ainsi pas nécessaire

**Sécurisé** — fonctions de sécurité de périmètre, d'authentification, d'autorisation granulaire et de protection de données (via chiffrement et gestion de clés) conformes aux réglementations

**Maîtrisé** — audit de données, traçabilité et découverte de données

**Administré** — une interface universelle fournissant des fonctions d'administration de bout en bout et d'autres fonctions critiques d'entreprise, telles que les mises à jour à chaud (sans interruption de service)

**Ouvert** — une plate-forme ouverte reposant sur l'open source et des standards ouverts pour que les entreprises puissent s'assurer que leur système Hadoop est durable, portable, mieux intégré dans l'écosystème et de très grande qualité.

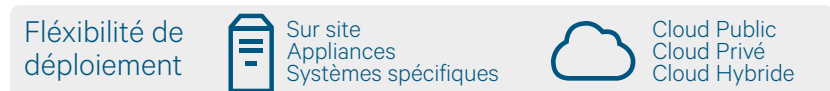
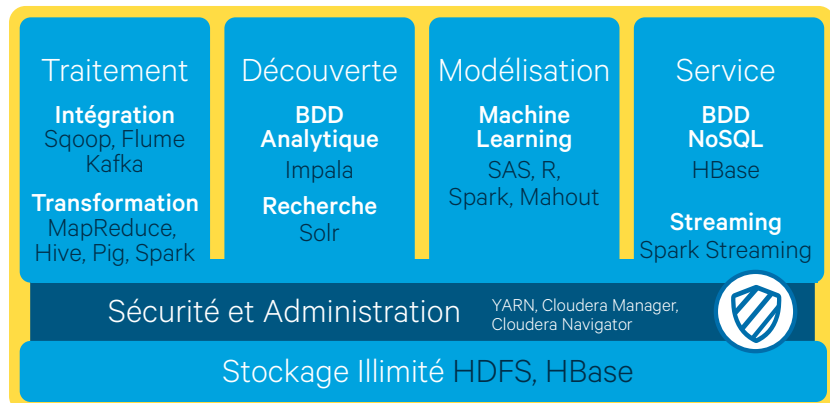


Illustration 1 : Cloudera Enterprise combine une distribution Hadoop 100% open source pour l'entreprise, des fonctions de gestion de données et de systèmes, et un support complet, au sein d'une puissante plate-forme unifiée de gestion de données.

**Véronique DAVAL**

Présidente du Club

+33 (0)1 53 45 28 65

club@decisiondsi.com

33 - rue Galilée 75116 Paris France



www.clubdecisiondsi.com



**CLUB DECISION DSI**  
1<sup>er</sup> Club Français de décideurs informatiques & télécoms

## LE CLUB DÉCISION DSI, PREMIER CLUB FRANÇAIS DÉDIÉ AUX DIRECTEURS INFORMATIQUES

Aujourd'hui le Club Décision DSI réunit 1 250 membres «directeurs Informatique & Télécoms» de sociétés du secteur privé et du secteur public dont le parc informatique est supérieur à 200 pc. Il rassemble des hommes et des femmes d'expérience, animés par la volonté de mener des actions qui soutiennent d'autres directeurs des systèmes d'information (DSI) dans la réalisation de leurs projets professionnels.

Le club est administré par Véronique Daval, la présidente, le vice-président, ainsi que par un conseil central composé de neuf directeurs informatiques.

### Chaque année le club organise :

- 20 manifestations thématiques par an.
- 7 événements accessibles aux partenaires (éditeurs de logiciels, constructeurs, opérateurs télécoms, SSII)
- Les Rencontres de l'innovation technologique
- Le consortium du DSI
- Les DSI Meeting (table rondes, débats)
- Le dîner de Gala avec son village solutions (réservé aux membres et aux partenaires du club)
- Le dîner de gala annuel

## HELP DSI , le 1er Réseau francophone de 1250 DSI connectés

HELP DSI ce sont des DSI, des groupes de DSI et des communautés de DSI qui échangent en associant leurs idées, leur expérience et leurs connaissances Afin d'optimiser leurs choix technologiques.

Le club dispose de son propre réseau social (RSE) sur lequel de nombreuses thématiques sont abordées qui permettent aux membres de capitaliser sur l'expérience des uns et des autres. Les ambassadeurs de Club Décision DSI ont adopté cette solution afin d'améliorer la communication entre les membres, de créer des groupes de réflexion sur des thématiques variées, et de diffuser les compétences à travers le réseau des membres au niveau national et européen.

Ce moteur d'expérience permet aux sociétés adhérentes de se servir des retours d'expérience des projets IT mis en place par les membres.

Chaque année les ambassadeurs identifient des thématiques auprès des membres et définissent un programme annuel consultable sur le site internet du Club.

## ORGANIGRAMME

### LES MEMBRES DU BUREAU ET AMBASSADEURS DU CLUB

**CLUB DECISION DSI**  
1<sup>er</sup> Club Français de décideurs informatiques & télécoms



**Véronique Daval**  
Présidente



**Julien Daval**  
Vice Président



**Armand ASSOULINE**  
CIO  
MSC France

**Gilles BERTHELOT**  
RSSI  
GROUPE SNCF

**Christian DOGUET**  
DSI  
CHAINE THERMALE DU SOLEIL

**Damien GRIESSINGER**  
CTO  
EPP0

**Christophe GUILLARME**  
DSI  
GROUPE AB

**Trieu HUYNH-THIEN**  
DSI Adjoint  
Centre Georges Pompidou

**Bertrand LENAIN**  
DSI  
CCI DE SEINE ET MARNE

**Lionel ROBIN**  
DSI  
GROUPE LA RESERVE

**Dominique TROUVE**  
DSI  
HOPITAUX AVICENNE

**Jiddou WAGHEF**  
DIR. DE L'ANIMATION DES SI  
CER FRANCE

**Claude YAMEOGO**  
ARCHITECT SI

CLUB FRANÇAIS DE 1200 DECIDEURS INFORMATIQUES



# Devenez Membre du **Club Décision DSI**

Le 1<sup>er</sup> Club Européen indépendant de Décideurs informatiques

*Décideurs informatique,  
élargissez votre cercle de  
relation, identifiez les solutions  
innovantes et tendances de  
demain, consolidez vos  
compétences*



**Club Decision  DSI**

16, Place Vendôme • 75001 Paris  
Tél. : 33 (0)1 53 45 28 65

[www.decisiondsi.com](http://www.decisiondsi.com)



COHERIS EST UN ÉDITEUR DE SOLUTIONS CRM ET ANALYTIQUES QUI PERMETTENT DE GÉRER LA RELATION CLIENT ET D'ANALYSER TOUTES LES DONNÉES DES ENTREPRISES POUR PILOTER LA PERFORMANCE ET TRAITER L'EXPLOSION DES DONNÉES NUMÉRIQUES.

DE LA PME À LA MULTINATIONALE, PLUS DE 1 000 ENTREPRISES ET COLLECTIVITÉS ONT DÉJÀ FAIT CONFIANCE À COHERIS DANS PLUS DE 80 PAYS. COHERIS S'APPUIE SUR UN RÉSEAU ÉTENDU DE PARTENAIRES INTÉGRATEURS ET CONSEIL, ET SUR SES PROPRES EXPERTS POUR OFFRIR À SES CLIENTS DES SOLUTIONS À LA FOIS OPÉRATIONNELLES, ANALYTIQUES ET PRÉDICTIVES AU SERVICE DE LEURS PERFORMANCES.

**Coheris Analytics Liberty** est une solution agile de **Business Intelligence** de gestion des données et de pilotage décisionnel permettant ainsi de traiter, explorer, analyser et restituer toutes les données de l'entreprise.

Intégrant **Dataviz**, **Data Storytelling** et **Analyse prédictive**, Coheris Liberty est une plateforme complète de développement d'applications analytiques qui permet à toute l'entreprise de profiter du potentiel de la Business intelligence.

Notre client IFOP a choisit la solution agile de Business Intelligence de Coheris pour l'analyse de ses études et la restitution dynamique des résultats.

« Après avoir étudié plusieurs solutions du marché, Coheris nous a convaincu par son expertise et sa solution très aboutie en matière de restitution de résultats. Nous avons besoin d'un acteur qui parle le même langage que nous et Coheris a nettement fait la différence. Lors du déploiement de la solution, nous avons apprécié la disponibilité des équipes, ainsi que l'accompagnement et la réactivité de Coheris » explique **Thomas Duhard, Directeur Production et Innovation chez IFOP.**

**Coheris Analytics SPAD** solution de référence pour gérer, explorer et modéliser les données, dédié au **data mining**, **text mining** et à **l'analyse prédictive**, capable d'exploiter des bases de données de toute dimension et de toute origine. Coheris SPAD propose plus de 70 méthodologies éprouvées et fiables. La solution est reconnue pour son ergonomie intuitive, la puissance de son moteur de calcul, sa facilité d'utilisation et la qualité graphique de représentation des résultats. Coheris SPAD permet, à partir des données de l'entreprise, d'anticiper les risques, d'identifier les opportunités, et d'optimiser les différentes opérations métier.

Notre client VEGA a choisit la solution de Dataming pour optimiser leur efficacité commerciale et marketing.

« Nous avons besoin de rationaliser les coûts et nous voulions des campagnes marketing plus efficaces en ciblant mieux les clients. Notre objectif était d'avoir la bonne offre, au bon moment, au bon client et via le bon canal. Nous avons fait appel à Coheris pour nous aider dans ce projet. Grâce à la solution Coheris Spad, nous avons pu traiter un très grand volume de données et avoir une meilleure connaissance de nos clients. Nous avons réduit nos coûts de 47% et augmenter le CA de la gamme « Hôtellerie » de 17% ». **explique Sylviane Lopez, DG Vega France.**

Nous avons développé des synergies entre nos solutions de gestion de la relation client et nos solutions analytiques, pour disposer de solutions innovantes complètes et agiles pour l'engagement client.

Coheris CRM suite est une solution globale de **gestion de la relation client**, qui vous permet de structurer et exploiter au mieux la **connaissance client** dans l'entreprise. Particulièrement puissante et robuste, elle dispose d'une approche analytique intégrée, de connecteurs aux réseaux sociaux, et permet un accès temps réel aux informations stratégiques. Cet outil **CRM** vous accompagnera dans la gestion de vos interactions clients omnicanal, afin d'accroître la performance commerciale de vos équipes, réduire vos coûts et fidéliser vos clients.

Notre client Crédit Agricole Titres choisit les solutions CRM et Business Intelligence de Coheris pour gérer la mobilité des clients des Caisses régionales.

« Lors des consultations, les solutions et les équipes de Coheris se sont nettement détachées des autres prestataires, avec une intégration très forte entre le CRM et la BI. La satisfaction des clients est au centre de l'organisation de Crédit Agricole Titres. Les solutions CRM et BI de Coheris nous permettent de répondre à cette exigence et de traiter le transfert de nos clients en cas de changement de domiciliation bancaire avec efficacité. La rapidité de mise en œuvre, la capacité d'intégration avec le SI du Groupe et la montée en compétences des équipes Crédit Agricole Titres nous ont confortés dans notre choix.» **explique Stéphane Brou, Directeur général de Crédit Agricole titre.**

Les solutions de **Customer Intelligence** de Coheris aident les entreprises dans la mise en place de leurs projets d'étude de Connaissance Client. Nos Datascientists accompagnent les fonctions métiers dans toutes les étapes nécessaires de leur projet pour transformer leurs données en capital stratégique.

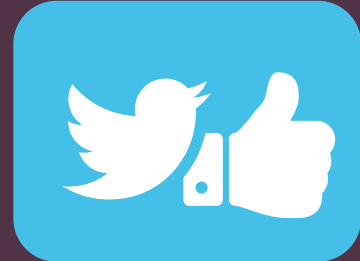
« L'exploitation des données est un atout majeur pour développer la valeur client. Nous avons fait appels aux équipes de Coheris pour nous accompagner dans notre projet de connaissance client, ce qui nous a permis d'augmenter notre CA, limiter les pertes et gagner en productivité. » **explique Cyprien Rouits, Responsable Pôle Décisionnel, Editions Atlas.**



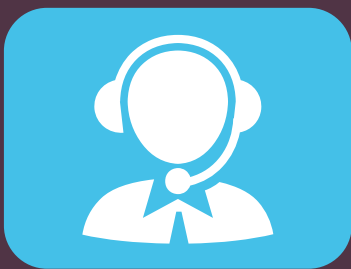
Retrouvez-nous sur  
[www.coheris.com](http://www.coheris.com)



STRATÉGIE  
DATA



SOCIAL CRM



SERVICE  
CLIENT



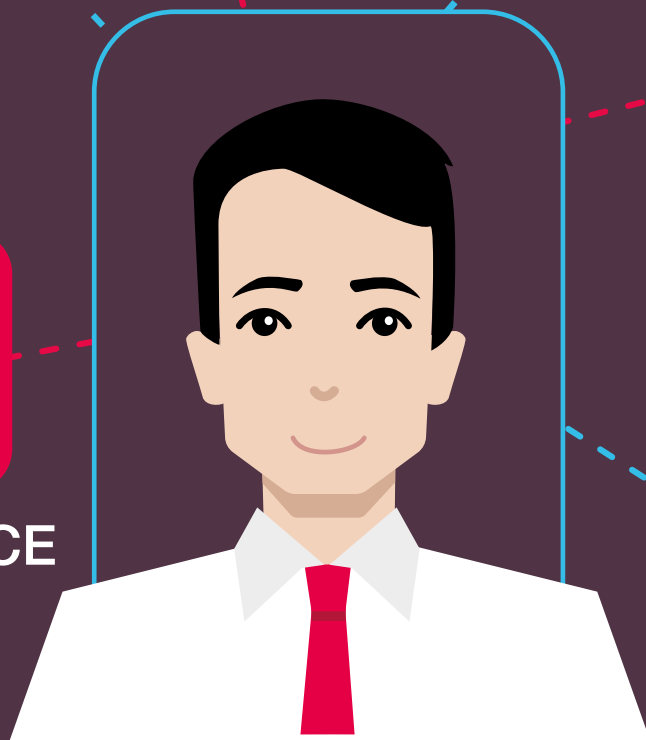
CONNAISSANCE  
CLIENT



PRÉDICTIF



CIBLAGE



 **COHERIS**

**Florence BONNET**

Présidente - Fondatrice

+33(0)1 46 98 90 01

info@cil-consulting.com

20 bis rue Louis Philippe 92200 Neuilly sur Seine France



## EN QUOI LE BIG DATA EST-IL IMPACTÉ PAR LA RÉGLEMENTATION SUR LA PROTECTION DES DONNÉES PERSONNELLES ?

Ce qui est nouveau avec les Big Data, c'est que le traitement porte sur un volume inédit de données d'une grande variété. Même si a priori il n'y a pas de données personnelles, il peut y avoir une possibilité de ré-identifier les personnes ou de faire du profilage. C'est aussi la multiplication des sources de risques liés aux solutions hébergées dans le nuage et/ou en dehors de l'UE. C'est enfin la complexité des traitements, généralement en temps réel ou quasi-réel, auxquels les mesures traditionnelles de sécurité ne sont pas adaptées.

Côté législation, les traitements de données personnelles doivent respecter un des principes « difficilement compatibles » avec les caractéristiques mêmes des Big Data, tels que la minimisation de la collecte, la proportionnalité du traitement, le principe de finalité compatible ou l'interdiction d'interconnecter des fichiers de finalités différentes. D'autre part dès le 25 mai 2018, le Règlement européen imposera aux entreprises de mener une étude d'impacts sur la vie privée pour les traitements présentant des risques élevés pour les personnes.

Par conséquent, avant la mise en œuvre d'un tel projet, les entreprises doivent s'interroger sur le cadre juridique applicable, sur la conformité et sur la sécurité des données. D'autant plus que la violation des dispositions du Règlement sera passible de sanctions de l'ordre de 4 % du C.A. mondial d'un groupe ou 20M d'euros.

## QUI EST CONCERNÉ ET QUELS SONT LES RISQUES EN CAS DE NON-CONFORMITÉ ?

La loi Informatique et Libertés concerne toutes les organisations, publiques ou privées, amenées à traiter d'informations sur les personnes, qu'il s'agisse de leurs salariés, de leurs clients, de leurs adhérents ou patients, des visiteurs d'un site internet ou des utilisateurs d'applications. En revanche, les mesures de sécurité doivent être adaptées aux risques liés aux traitements. Par conséquent, les organisations ayant à traiter un grand nombre de données sensibles ou mettant en œuvre des traitements susceptibles de présenter des risques pour les libertés et droits fondamentaux des personnes sont plus particulièrement concernées quelle que soit leur taille.

Aujourd'hui la loi concerne uniquement les responsables de traitement, tandis que les sous-traitants ont une responsabilité contractuelle à leur égard.

Surtout, dès mai 2018, les sous-traitants seront soumis aux mêmes obligations de sécurité et passibles de sanctions.

## QUELLE EST LA DÉMARCHÉ À SUIVRE PAR LES ORGANISATIONS SOUHAITANT METTRE EN PLACE DES PROJETS DE BIG DATA ?

Cette problématique doit être intégrée dès le lancement et tout au long d'un tel projet.

En premier lieu, l'organisation devrait désigner un responsable de la protection des données, un Correspondant Informatique et Libertés par exemple, et l'intégrer à l'équipe projet. Il aura un rôle d'interface et de facilitateur notamment entre les juristes, le responsable de la sécurité et les métiers. Sa mission sera d'analyser la conformité de la collecte et des traitements ultérieurs des données, de sensibiliser l'équipe aux principes de protection des données, de documenter les traitements, de conseiller le porteur de projet sur la nécessité de mener une étude d'impacts, de proposer une démarche de « Privacy by Design ». L'objectif est bien d'éviter une remise en cause ultérieure du projet qui serait contraire à la loi puisque la CNIL a le pouvoir d'interdire les traitements et de prononcer des sanctions qui peuvent être rendues publiques.

La démarche peut être partiellement industrialisée sur la base de cas d'usages. La protection des données personnelles, y compris la sécurité, doivent être intégrées aux projets. La matière est complexe ; aussi il est conseillé de recourir à des professionnels compétents et expérimentés. Vous dites que la Privacy By Design doit permettre d'intégrer les principes de protection des données personnelles dans les Big Data. Comment faire ?

Face à l'évolution des technologies, la loi ne suffit plus ; il faut d'autres modes de régulation pour garantir la protection des données et de la vie privée. La Privacy by Design doit permettre de renforcer la protection des personnes dont les données sont traitées.

Il n'existe pas encore de standard de Privacy By Design ; cette approche trouve son origine au Canada où elle a été développée par Ann Cavoukian alors Commissaire à la protection de la vie privée de la province de l'Ontario.

Aujourd'hui cela doit s'entendre comme une démarche globale, à la fois organisationnelle, juridique et technique visant à intégrer les principes de protection des données tels que la minimisation, la transparence ou la sécurité notamment sous forme d'anonymisation, de pseudonymisation et de chiffrement. Il s'agit aussi de redonner la maîtrise aux personnes sur les données les concernant et d'utiliser des PET's (Privacy Enhancing Technologies).

Le Règlement applicable en mai 2018 consacre la Privacy by Design et précise que le responsable du traitement doit protéger les données dès la conception et par défaut. Lors de l'élaboration, de la conception, de la sélection et de l'utilisation d'applications, de services et de produits il est dit qu'il convient d'inciter les fabricants et les prestataires de services à prendre en compte le droit à la protection des données lors de l'élaboration et de la conception de tels produits, services et applications et de s'assurer que les sous-traitants soient en mesure de s'acquitter des obligations qui leur incombent.

Le Règlement ajoute que ces principes devraient être pris en considération dans le cadre des marchés publics.

Les exigences doivent aussi être traduites en langage informatique, juridique et en politiques et procédures.



Protection des Données

---

# Société de conseil en protection des données personnelles

---



Conformité & Sécurité des données  
Audit CNIL - Formation - Privacy Certification  
Correspondant Informatique et Libertés (CIL) /  
Data Protection Officer (DPO)



## CONTACTEZ NOUS :

✉ **COMPLIANCE IT LEGAL CONSULTING**  
20 bis rue Louis Philippe - 92200 Neuilly/Seine  
Tel. 01 46 98 90 01 - info@cil-consulting.com

🐦 @CILCONSULTING

[www.protection-des-donnees.fr](http://www.protection-des-donnees.fr)

# CONVERTEO

Adriano MUCCIARDI

Manager

+33(0)6 66 79 38 26

+33(0)1 84 17 26 21

amu@converteo.com

15 Place de la Nation 75011 Paris France

CONTACT



www.converteo.com



SMART DIGITAL & DATA CONSULTANTS



## LES DATA-LAKES AU SERVICE DU SMART MARKETING ? LE POINT DE VUE DE CONVERTEO

Dans un marketing de plus en plus « dataifié », ces dernières années ont été marquées par la multiplication des technologies permettant de centraliser et activer la data marketing.

La dernière en date est le data-lake, outil adopté par certaines entreprises, notamment afin d'améliorer leur efficacité marketing et commerciale.

<p><b>La SNCF</b></p> <p><b>Objectif :</b> Prévenir et réduire l'usure des équipements</p> <p><b>Méthode :</b> adoption d'un data-lake on premise avec cluster hadoop pour faire des analyses croisées (identifier les zones à risque de gel sur les rails, de développement de végétations sur les voies...) en installant de nouvelles rampes connectées et en intégrant des données topographiques issues de drones</p> <p><b>Résultat visé :</b> Meilleur service et moindre usure du matériel</p>	<p><b>Oscaro (spécialiste des pièces détachées automobiles)</b></p> <p><b>Objectif :</b> croisement des données de navigation du site e-commerce avec le catalogue produits (et d'autres sources)</p> <p><b>Méthode :</b> adoption d'un cluster Hadoop interne de 10 serveurs, alimenté par 80 sources de données • BigQuery pour traitement • BI pour analyse et visualisation</p> <p><b>Résultat visé :</b> Permettre aux services vente et marketing de produire des analyses détaillées agiles et facilement</p>	<p><b>Uber</b></p> <p><b>Objectif :</b> Mettre au point une tarification en temps réel pour les chauffeurs</p> <p><b>Méthode :</b> Utilisation de Spark (gestion des flux de données) et d'Hadoop (HDFS) pour développer un algorithme prédictif de l'offre et de la demande en temps réel (en prenant en compte le jour, la période de la journée, les conditions météorologiques...) du prix de la course, du meilleur trajet à suivre et des potentiels trajets</p> <p><b>Résultat :</b> Estimation de la demande 5 min à l'avance et détermination du prix exact de la course</p>	<p><b>Walmart</b></p> <p><b>Objectif :</b> Avoir une vision à 360° des 250 millions de clients hebdomadaires et de leurs besoins pour gérer les 11 000 magasins dans le monde</p> <p><b>Méthode :</b> Plateforme big data qui regroupe des données issues de 200 sources internes externes (classement des ventes, inventaires produit, plong, conditions climatiques, état de l'économie...) et prédit les ventes et la demande par magasin, mais aussi les localisations des futures enseignes</p> <p><b>Résultat :</b> Amélioration des processus opérationnels (supply chain, marketing...)</p>
--	--	---	---

Revenons sur cette technologie qui, bien exploitée, peut devenir la pierre angulaire de la stratégie data et marketing d'une organisation.

S'appuyant majoritairement sur une infrastructure dite « big data », le data-lake serait un datawarehouse dont on aurait gommé la plupart des limites :

- Potentiel de stockage quasi illimité
- Puissance de calcul quasi illimitée
- Capacité à se connecter à l'ensemble des systèmes d'information et des applicatifs d'une organisation
- Compatibilité avec des données structurées et non structurées

Contrairement au datawarehouse, qui a uniquement vocation à stocker les données « utiles » et principalement structurées, le data-lake stocke toutes les données, qu'importe leur format, sans schéma de données préconçu. Il possède une architecture flexible et ouverte, alors que le datawarehouse est un espace de stockage et de restitution structuré et relativement figé.

## LE DATA-LAKE COMME MOYEN DE TRANSFORMER L'ENTREPRISE

Les données issues de l'ensemble de l'entreprise se retrouvent ainsi au même endroit, ce qui facilite donc l'accès, l'analyse et l'exploitation de ce patrimoine data ainsi constitué. Les data-lakes sont un des agents du « désilotage des organisations ».

La grande accessibilité de cette donnée couplée avec une grande puissance de calcul permet de démocratiser l'utilisation avancée de la donnée. Il y a encore peu, chaque analyse d'un jeu de données issu de plusieurs systèmes, se traduisait par un projet structurant et engendrant des délais conséquents sans certitude sur le résultat. Les équipes métier se retrouvaient souvent à s'autocensurer et vivre de vraies frustrations, alors que la donnée existait et semblait riche en potentiel.

Outil de démocratisation de la donnée, le data-lake permet aujourd'hui aux équipes métier d'amorcer une démarche data-driven ; où la simplicité et la rapidité aident à redévelopper la curiosité du chiffre.

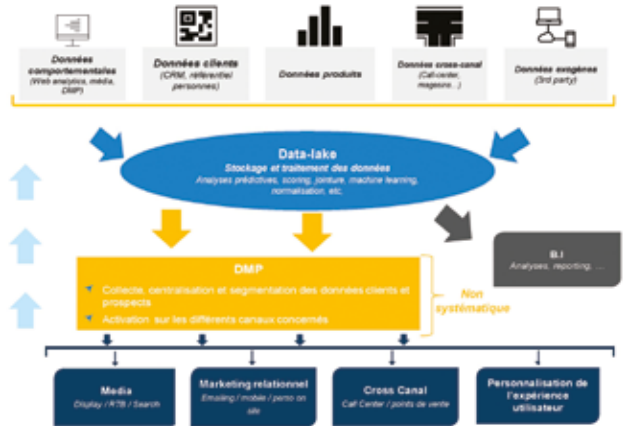
Loin des contraintes techniques du passé, les limitations volent en éclat, les équipes peuvent enfin se concentrer sur la pertinence de leurs analyses.

## FOCUS SUR LES COMPLÉMENTARITÉS ENTRE DMP (DATA MANAGEMENT PLATFORM) ET DATA-LAKE

Deux principaux points prouvent leur forte complémentarité :

- A la différence du data-lake, la DMP n'a pas vocation première à stocker les données personnelles (PII) et autres données sensibles de l'entreprise
- La DMP est directement reliée à l'écosystème digital (de préférence en temps réel), tandis que le data-lake est davantage adapté à un fonctionnement d'imports/exports quotidiens de données.

DMP et data-lake sont donc à envisager comme deux « briques » qui se superposent : le data-lake agrège et calcule la donnée potentiellement sensible, via des opérations telles que le scoring ou le calcul prédictif, avant de la transmettre à la DMP (sous forme anonymisée et simplifiée), pour que celle-ci l'exploite dans sa segmentation et sur les différents canaux auxquels elle est connectée.



Deux modes d'hébergement possibles : On Premise et Cloud Services. Avec le mode «On Premise», l'organisation a la mainmise sur l'ensemble de la chaîne de stockage et d'exploitation de la donnée. L'organisation doit également fixer ses propres règles de sécurité à appliquer à sa donnée. Les architectures Cloud présentent, elles, de grands avantages en termes de facilité de déploiement et de ressources nécessaires à l'exploitation des données, en mettant à disposition des entreprises des produits de plus en plus packagés et automatisés ; selon une philosophie « NoOps ».

L'entreprise doit choisir le mode d'hébergement qui répondra le mieux à ses besoins en termes de gouvernance, d'évolution et de facilité de déploiement. Les deux modes d'hébergement présentant chacun des avantages spécifiques.

## QUELLE ORGANISATION ET EXPERTISE POUR LE LANCEMENT D'UN PROJET DATA-LAKE ?

La complexité de mise en place d'un data-lake demande la constitution d'une équipe projet ad hoc, bénéficiant d'un fort sponsorship interne (la donnée étant transversale à toute l'entreprise).

Cette équipe doit idéalement présenter un administrateur système, un data architecte, un data engineer ainsi qu'un data scientist. Enfin, une méthode agile est à privilégier, afin de garantir un déploiement progressif, au rythme des différents cas d'application préalablement priorisés en fonction de leur intérêt.





### SMART MARKETING

- Parcours client cross-canal
- Media & audience management
- Offre produits et services

### SMART DATA

- Collecte et ingénierie
- Analytics et Data Science
- Activation et automatisation

### SMART ORGANISATION

- Expertise et culture digital & data
- Conduite agile du changement
- Gouvernance et processus

Expert  
 Indépendant  
 Stratégique  
 &  
 Opérationnel

50 Consultants

### Plus de 100 clients depuis 2007



### Parmi nos partenaires technologiques



Seul acteur français avec la triple certification Google : GA360 / Doubleclick – Programmatic Consulting / Cloud Platform



# Couchbase

## AVEC COUCHBASE SERVER, CRITEO STIMULE LA PERFORMANCE ET LA MONTÉE EN CHARGE HORIZONTALE DE LA PLATE-FORME PUBLICITAIRE NUMÉRIQUE

Criteo, l'un des géants de la publicité numérique, cible chaque mois plus d'un milliard d'utilisateurs Internet uniques du monde entier. Son moteur automatise la recommandation de publicités et la sélection de produits des catalogues des annonceurs – 30 milliards de fois par jour, en moins de 100 millisecondes et après un calcul faisant intervenir plus de 100 variables.

Avec un tel volume, la performance et la montée en charge horizontale sont essentielles pour proposer au moment opportun l'annonce adéquate au bon utilisateur. Criteo a testé différentes bases de données et a retenu Couchbase Server pour leur cas d'utilisation en raison de l'aptitude sans égal de Couchbase à gérer un nombre élevé de requêtes et à offrir des temps de latence extrêmement faibles.

### UNE RÉPONSE IMMÉDIATE AUX ENCHÈRES EN TEMPS RÉEL

Criteo utilise des algorithmes prédictifs et de nombreuses données sur les intentions et les habitudes d'achat des consommateurs pour tarifier et fournir des annonces personnalisées. « Pour offrir la bonne annonce personnalisée, nous avons besoin de solutions techniques qui nous permettent de cibler les visiteurs en temps réel et d'exploiter précisément des données complexes », explique Nicolas Helleringer, directeur de l'ingénierie en abilité de site chez Criteo. « Nous avons besoin d'un accès en temps réel à un grand nombre de données utilisateur, que nous ne pouvons obtenir avec des bases de données relationnelles. »

La société, qui a servi 741 milliards d'annonces en 2014, a essayé plusieurs autres bases de données NoSQL avant d'opter pour Couchbase. « Nous en sommes arrivés à la conclusion que Couchbase Server était parfaitement adapté pour fournir la montée en charge horizontale et la performance indispensables dans certains secteurs de données », précise Romain Niccoli, co-fondateur de Criteo et directeur des techniques informatiques.

« Couchbase Server nous permet de gérer la charge et de répondre à la nécessité absolue consistant à répondre en temps opportun aux requêtes d'enchères en temps réel (RTB) », ajoute Nicolas Helleringer.

« Aucune autre solution SQL ou NoSQL ne pouvait assumer ainsi cette tâche. » La fiabilité et la facilité de montée en charge de Couchbase Server suivant notre croissance étaient précisément ce dont nous avions besoin. Et ce sans cesser de répondre aux requêtes RTB en quelques millisecondes – ce qui est vraiment extrêmement rapide. »

### UNE TRANSITION EN DOUCEUR SUR PLUS DE 1 000 SERVEURS

En 2011, Criteo a déployé Couchbase Server sans aucun problème. « Un déploiement sans à-coups eu égard à la performance et à la montée en charge horizontale », indique Nicolas Helleringer. « Nous avons d'abord utilisé Couchbase Server sur un petit projet. De mois en mois, son utilisation a augmenté, car la solution autorisait la performance que nous recherchions. Elle s'est même avérée très, très productive. »

L'installation de Couchbase Server par le géant mondial du service de publicité a connu un développement extrême, pour totaliser plus de 1 000 serveurs répartis en 24 grappes, soit 107 téraoctets de RAM et d'espace sur disque statique (SSD). « Nous recourons à l'automatisation pour le déploiement sur des serveurs en métal nu », explique Nicolas Helleringer. « Couchbase Server fait partie de cette infrastructure. Nous avons réalisé le déploiement sur des centaines de serveurs avec Chef, un outil d'automatisation du processus de dimensionnement et de configuration des serveurs et des logiciels qu'ils exécutent. »

Couchbase Server prend également en charge la réplication croisée entre centres de données (XDCR), ce qui implique dans Couchbase Server la réplication de données actives vers de multiples centres de données, de localisation géographique diverse. L'objectif est alors la reprise en cas de sinistre ou la mise à disposition des données plus près des utilisateurs afin d'accélérer l'accès aux données. Criteo s'appuie sur XDCR pour répliquer de façonnable les données vers sept centres de données du monde entier. Les données sont alors plus proches des utilisateurs, ce qui permet à Criteo de fournir plus rapidement encore les annonces adaptées.

« Pour nos exigences de charge et de performance – et nous sommes loin d'être la seule société exigeant un débit élevé et des demandes à faible latence – Couchbase Server est la solution idéale, qui allie toute la puissance dont nous avons besoin à une extrême simplicité d'utilisation ». Nicolas Helleringer, Directeur de site Reliability Engineering.

### RETOUR SUR INVESTISSEMENT ET AVANTAGE CONCURRENTIEL

Pour Criteo, le retour sur investissement de Couchbase Server est déjà effectif. « Avec 73 % de croissance d'une année sur l'autre, nous allons poursuivre le déploiement de Couchbase Server pour assurer les temps de réponse de l'ordre de la milliseconde dont dépendent nos annonceurs pour atteindre leurs clients cible », précise Romain Niccoli.

Les outils de rapport de Couchbase Server ont considérablement réduit la durée et les efforts indispensables pour que Criteo gère son déploiement, ce qui nous permet d'utiliser notre personnel technique à des fins plus stratégiques. Nicolas Helleringer ajoute alors : « La technologie nous permet de conserver l'avance que nous avons sur la concurrence. Sans compter que sa gestion est simple et que les dépenses administratives associées sont faibles, même pour une grande structure telle que la nôtre. »

### UNE PERFORMANCE HOMOGÈNE, SANS TEMPS D'ARRÊT

Parmi les bases de données NoSQL testées précédemment par la société, beaucoup étaient « vraiment délicates et complexes à utiliser », affirme Nicolas Helleringer. Avec Couchbase Server, « Nous disposons d'une base de données NoSQL fiable, qui fournit une performance homogène sans problème ni temps d'arrêt. »

### LA SOLUTION NOSQL LA PLUS EFFICACE

« Pour nos exigences de charge et de performance – et nous sommes loin d'être la seule société exigeant un débit élevé et des demandes à faible latence – Couchbase Server est la solution idéale, qui allie toute la puissance dont nous avons besoin à une extrême simplicité d'utilisation », poursuit Nicolas Helleringer.



# LA BASE DE DONNÉES, RÉINVENTÉE

## COUCHBASE SERVER 4.5

La base de données la plus  
puissante et complète.

L'évolutivité du NoSQL. La flexibilité du JSON.  
La puissance du SQL.



## QUELLE EST L'ADN DE COUTHON CONSEIL ?

Couthon Conseil est un cabinet de conseil, spécialisé dans le **recrutement par approche directe** et le **coaching de carrière** de talents de la Data (*Big Data, Data Science, Analytics, IoT, DataViz, Open Data, etc.*) et du Digital (*IT, Transformation Digitale, Webmarketing, UX/UI Design, etc.*).

Ayant tous exercé dans les métiers qu'ils adressent au quotidien, les collaborateurs de notre cabinet sont tout autant chasseurs de têtes pour nos clients qu'agents de carrière pour nos candidats.

Cette dualité de rôles complémentaires pérennise efficacement nos relations clients/candidats et nous permet d'aborder nos missions de manière globale et pragmatique.



## NOTRE POSITIONNEMENT DISRUPTIF

Couthon Conseil dépasse les codes du recrutement tel qu'il est encore (*trop souvent*) pratiqué et ainsi perçu par les candidats et les entreprises, c'est-à-dire (*trop souvent*) industriel et guidé uniquement par une culture du chiffre, au détriment de la qualité des prestations et des relations humaines.

Le candidat est au cœur de nos préoccupations. C'est d'autant plus crucial sur un marché en plein essor comme celui de la Data, où les besoins en recrutement sont considérables et les talents pénuriques.

Une connaissance fine des métiers de la Data et un ancrage fort dans cette communauté (*dynamique partenariale, interventions dans des formations spécialisées, participation à des événements, etc.*) s'avèrent indispensables si l'on souhaite en saisir pleinement les enjeux et incarner un véritable rôle de Conseil en phase avec la réalité opérationnelle des métiers sur lesquels nous travaillons.

Couthon Conseil a développé toute une gamme de services **par et pour des opérationnels**.



C'est d'ailleurs dans cet esprit que nous éditons le **D²Blog** (*Data x Digital Blog*), blog où les contributeurs sont des professionnels français et internationaux de la Data et du Digital.

## POUR NOS CLIENTS : CHASSE DE TÊTES ET RECRUTEMENT

Pionnier en France dans la chasse de têtes et le recrutement de talents dans la Data, Couthon Conseil propose à ses clients une approche holistique et disruptive du recrutement, basée sur une démarche de conseil qualitative et des expériences opérationnelles réussies.

En France, comme à l'international, nous adressons :

- **tous niveaux d'expérience** (*du stagiaire au directeur*)
- pour **toutes fonctions dans la Data** (*scientifiques, techniques, commerciales et managériales*)
- dans **tous secteurs d'activité** (*services, industrie, banque, etc.*)
- et **toutes tailles d'entreprise** (*startups, PME, grands comptes, cabinets de conseil et ESN*)

### Votre dénicheur de talents dans la Data



## POUR NOS CANDIDATS : COACHING DE CARRIÈRE

Une carrière ne se bâtit pas uniquement au moment des changements de poste, mais sur la durée.

C'est pourquoi nous offrons à nos candidats un accompagnement complet :

- conseils en **gestion de carrière**,
- développement du **personal branding**,
- accès à des **opportunités exclusives** cachées du marché,
- **coaching**,
- etc.

### Votre agent de carrière dans la Data



## CONSEIL & SERVICES COMPLÉMENTAIRES

En collaboration étroite avec nos experts, nous réalisons des missions de conseil (*audits Data, mise en place de stratégies Data, accompagnement à la mise en place et développement d'équipes Data, etc.*).

Couthon Conseil est également le créateur du premier service d'**évaluation personnalisée et gratuite de votre rémunération** dans la Data : [www.MonSalaireDansLaData.fr](http://www.MonSalaireDansLaData.fr)



**monSalaireDansLaData.fr**  
by COUTHON CONSEIL

# Les métiers porteurs dans la Data

by COUTHONCONSEIL®

## BIG DATA ARCHITECT

Expert des infrastructures IT permettant le stockage, la manipulation et la restitution des « méga données », il conçoit et administre des Data Centers, en hybride ou dans le cloud sur des plateformes comme Amazon AWS ou Microsoft Azure. Il travaille en amont dans la chaîne de traitement de la donnée et est le pilier de tout projet Big Data.



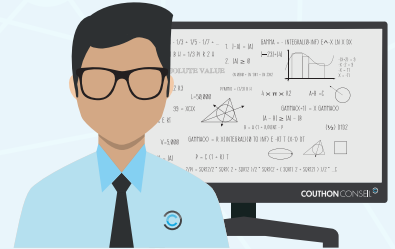
## DATA ANALYST

Plutôt en fin de chaîne des projets Data et avec l'appui du Data Scientist sur les dimensions technico-scientifiques, il se concentre sur l'exploration et l'exploitation des données métier, dont il extrait des KPI pertinents. Il peut ainsi vulgariser et restituer les résultats aux décideurs, notamment avec des Data Visualisations.



## DATA SCIENTIST

Au cœur des projets Data, il s'appuie sur ses compétences techniques et scientifiques avancées, contextualisées par des connaissances métier indispensables. Il élabore des algorithmes complexes, utilise des outils mathématiques, statistiques et du marché (SAS, SPSS, R, etc.) pour extraire, analyser et transformer des données (massives ou non) en information pour répondre à un besoin métier.



## DATA CONSULTANT

Interagissant avec les divers acteurs de la chaîne des projets Data (du Big Data Architect au Chief Data Officer), il aide les entreprises à définir et à implémenter leurs stratégies Data. Sa connaissance générale des outils du marché, sa créativité et sa compréhension des enjeux métier lui permettent de leur proposer des solutions innovantes.



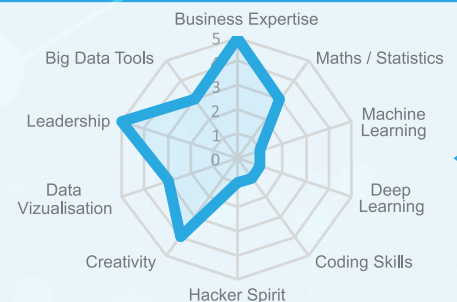
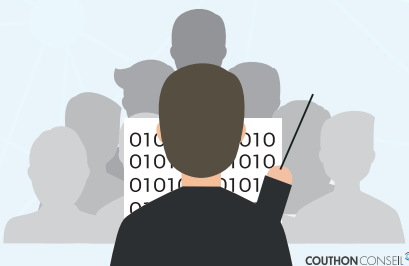
## CHIEF TECHNOLOGY OFFICER

Manager de haut niveau expérimenté rattaché à la Direction Générale, il est en charge des outils et des solutions technologiques innovantes, dont il est l'instigateur au sein de l'entreprise. Il pilote leur conception, leur mise en œuvre et les fait évoluer.



## CHIEF DATA OFFICER

Manager de haut niveau expérimenté rattaché à la Direction Générale, il est responsable de toute la gouvernance des données et de leur valorisation. Il est le garant des données, de leur agrégation et de leur exploitation pour répondre aux enjeux décisionnels de l'entreprise.



Infographie réalisée par COUTHONCONSEIL en partenariat avec Data ScienceTech Institute, la 1<sup>re</sup> école supérieure d'application à la Data Science et au Big Data en France.

**Steve VAN DEN BERG**

Regional Vice President - Western Europe

+33 (0) 1 72 71 86 86

svandenber@datastax.com

121 avenue des Champs Elysées, 75008 Paris



www.datastax.com



## EDELIA, FILIALE À 100% DU GROUPE EDF, UTILISE APACHE CASSANDRA ET DATASTAX ENTERPRISE

**Grâce à DataStax et au système de gestion de base de données Apache Cassandra, les clients EDF ont la possibilité avec la solution e.quilibre de reprendre en main leur consommation énergétique et d'être acteur du bien-être durable de leur habitat.**

EDELIA, filiale 100% EDF, évolue dans le secteur de l'énergie. Elle a pour rôle de concevoir, développer et de mettre en œuvre des solutions de suivi et de mesure des consommations d'électricité et de gaz des logements collectifs et individuels.

EDELIA produit par exemple pour EDF une solution nommée e.quilibre. Cette solution permet notamment de suivre sa consommation d'électricité et de gaz en kWh et en € sur une base estimée, et de la comparer d'une année à l'autre.

### Les besoins d'e.quilibre : un outil de stockage et de traitement de big data performant

E.quilibre doit assurer le traitement des données transmises par EDF concernant ses clients particuliers (profils, index de consommations, éléments de comparaisons) ainsi que les données collectées progressivement par la nouvelle génération de compteur communicant d'ERDF actuellement en déploiement, « LINKY ». **Laurent Pelletier, directeur de l'ingénierie d'EDELIA explique ainsi :**

« Avec un nombre de données aussi important que ce que traite EDF, il est très difficile de trouver un système adapté. Nous nous sommes tournés vers une solution big data capable d'être performante et de soutenir un service web et une application grand public. Dans un premier temps, une étude comparative du marché a été menée auprès de différents acteurs. Apache Cassandra a été retenue pour sa versatilité et sa simplicité d'exploitation. De plus, sa scalabilité horizontale très simple, sa très haute disponibilité et son modèle de données en « Colonne » sont très adaptés au stockage de séries temporelles. Par ailleurs, Edelia a décidé de se tourner vers la version DataStax de Cassandra qui amène une plus grande sécurité et permet de bénéficier d'expertises et d'accompagnement. »

« La performance, la simplicité d'exploitation ainsi que le coût réduit grâce à l'open source ont convaincu Edelia », se réjouit **Steve Van Der Berg, Vice President Europe de l'Ouest chez DataStax**. « Par ailleurs, la solution DataStax les a séduits par le confort, la sécurité du support et le bon fonctionnement que DataStax leur garantissait. Le déploiement de cette base de données est le cœur de ce système et une technologie nouvelle. », conclut-il.

**Les bénéfices de DataStax : Un confort, une sécurité mais aussi un accompagnement technique**

Utilisateur de DataStax depuis l'été dernier, EDELIA se positionne comme précurseur dans le domaine du big data.

« En six mois d'utilisation, nous avons pu constater la rapidité et la robustesse de la solution Cassandra. DataStax nous a alimentés en conseils et propositions d'implémentations, que nous avons décidé de retenir ou non, et a pu nous mettre en relation avec d'autres industriels disposants de témoignages utilisateurs intéressants. La performance en termes de lecture et d'écriture est un élément différenciant par rapport aux autres outils que nous avons évalués. » commente **Michel Henry, directeur de la production chez EDELIA. Laurent Pelletier ajoute :** « Nous avons commencé petit, par des POC, et ensuite nous avons réalisé nos premières expérimentations avec Cassandra. Ces solutions big data sont nouvelles et représentent un certain challenge pour les développeurs, car il faut apprendre à réfléchir différemment. Ensuite nous sommes passés au mass market avec e.quilibre, qui tourne sur Cassandra. Nous constatons depuis l'utilisation de cette solution de bonnes performances permettant d'atteindre nos objectifs et de tenir les délais en termes de déploiement et capacité ».

Edelia constate donc une bonne performance et un bon accompagnement de DataStax dans le déploiement de la solution. **Michel Henry** voit plusieurs autres avantages à la solution Cassandra DataStax et notamment :

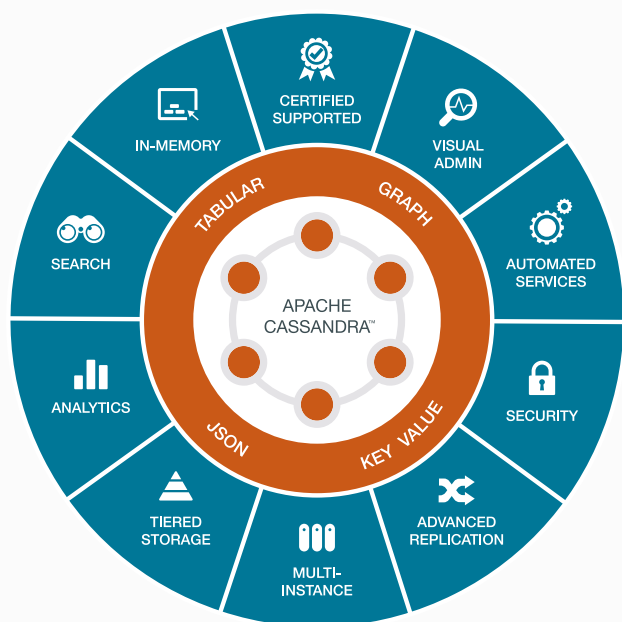
- **La Scalabilité horizontale très simple**, permettant d'ajouter des nœuds physiques au cluster en live
- **Sa très haute disponibilité**
- **Son modèle de données « colonne »**, très adapté au stockage de série temporelle
- **L'accompagnement et la formation** fournis par DataStax

### Une solution big data évolutive et adaptée dans divers domaines

« EDELIA innove techniquement en utilisant une solution big data telle que Cassandra DataStax. C'est une innovation technique au service d'une solution innovante qui est e.quilibre. » témoigne Michel Henry. Pour le moment, la plateforme EDELIA est composée d'un data center professionnel de 200 serveurs, Firewall et routeurs. Cinq nœuds Cassandra sont en production depuis l'été 2015.

### À propos de DataStax

DataStax a été créée en 2010 et se fonde sur le système de gestion de bases de données open source Apache Cassandra et cherche à faire face aux limitations des bases de données relationnelles. Le portefeuille client de DataStax comprend aujourd'hui plus de 500 entreprises dans 55 pays dont 30% appartiennent au Fortune 100. Plus de 400 employés travaillent aujourd'hui pour DataStax dans le monde dont la France où la société jouit d'une croissance rapide.



**DATASTAX ENTERPRISE**

## DataStax, leader des systèmes de gestion de base de données distribuées.

Système de base de données distribuées développé autour de l'open source Apache Cassandra™. Spécialement conçu pour les applications Internet, mobiles et objets connectés. Plus de 500 clients dans 50 pays, dont de grands noms comme Adobe, eBay, Intuit et Netflix.

- Technologie de base de données distribuées NoSQL
- Architecture continuellement disponible
- Extensibilité (scalabilité) prédictive et linéaire
- Big Data
- Sécurité d'entreprise (PCI DSS, LDAP, ...)
- Analytiques temps réel
- Stockage en mémoire ('In-memory')
- Déploiement multi-site et/ou Cloud
- Recherche avancée (géospatial, plein texte, ...)



DataStax France  
121 avenue des Champs Elysées  
75008 Paris  
Tel. +33 (0) 1 72 71 86 86

Contact : Steve van den Berg  
Regional Vice President  
Western Europe  
svandenberg@datastax.com

**Agile, continuellement disponible et extensible à volonté de façon prédictible.**

Pour plus d'informations visitez [www.datastax.com](http://www.datastax.com) ou suivez-nous sur Twitter: @DataStax



### SUCCESS STORY DENODO DATA VIRTUALIZATION AVEC AUTODESK

Autodesk est un leader dans la conception 3D, d'ingénierie et de logiciels de divertissement. Depuis son introduction du logiciel AutoCAD en 1982, Autodesk continue à développer le plus vaste portefeuille de logiciels 3D pour les marchés mondiaux. Les clients à travers la fabrication, l'architecture, la construction et les médias et les industries du divertissement, utilisent les logiciels Autodesk pour concevoir, visualiser et simuler leurs idées avant qu'elles ne soient construites ou créées. Compte tenu de l'évolution des modes de consommation du logiciel, la société a décidé de transformer son modèle d'affaires d'une licence perpétuelle classique à un modèle plus moderne de licence par abonnement pour augmenter les profits et propulser la croissance.

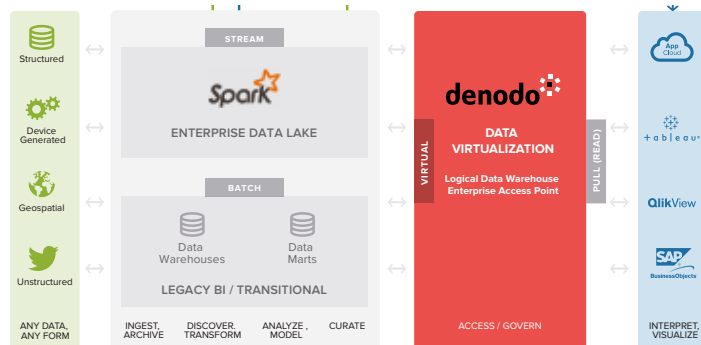
### BESOIN D'AFFAIRES

Le système existant Business Intelligence (BI) d'Autodesk ne pouvait pas soutenir ce changement crucial pour le modèle de revenus. La transition a affecté la capacité de la section des finances pour suivre les abonnements, les renouvellements et les paiements, et le système de BI, qui comprenait un entrepôt de données opérationnelles, qui ne pouvait pas répondre aux demandes des parties prenantes des entreprises, qui de plus en plus nécessitent à la fois des données de haute qualité et en temps réel. Autodesk a rapidement décidé que l'évolution vers une architecture agile BI 2.0 était nécessaire avec un entrepôt de données logique à sa base. Plus précisément, cela impliquerait le déplacement de leur modèle physique existant pour une approche plus logique d'intégration des données. Suite à la décision, un entrepôt de données logique en utilisant la plate-forme Denodo pour la virtualisation de données a été mis en œuvre pour répondre à ce changement.

### LA SOLUTION: ENTREPOT DONNEES LOGIQUE AVEC BIG DATA

Chez Autodesk, les données proviennent de multiples sources disparates de données inclus OLTP, fichiers plats, des données géo-spatiales, des flux de données de médias sociaux, et les journaux Web. Les systèmes d'intégration traditionnels inclus ETL pour le traitement par lots de données provenant de bases de données relationnelles et Kafka (CSE) pour l'intégration des flux de données. En outre, Autodesk avait un lac de données d'entreprise, qui a logé des grandes données de plusieurs sources de données telles que Spark et Scala, tandis que le OLTP et les données géo-spatiales étaient ETL-d et stockées dans un entrepôt de données opérationnelles. L'infrastructure a été lente et inefficace, et elle a été incapable de répondre aux demandes de renseignements d'affaires et l'agilité nécessaire pour répondre à leur transition vers un nouveau modèle de licence. La plate-forme Denodo pour la virtualisation de données a été introduite au service des finances d'Autodesk, qui permet de suivre les abonnements, les renouvellements et les paiements. La virtualisation des données a permis d'abstraire des données financières de leurs utilisateurs professionnels sans devoir déplacer ou à transformer physiquement des données. En outre, Autodesk a réalisé la nécessité de protéger l'ensemble des données sensibles appartenant à l'entreprise. Il y a actuellement plusieurs gouvernances, le risque et la conformité des initiatives (GRC) conçues pour aider Autodesk à mieux gérer le traitement des données sensibles, ainsi que pour atténuer les risques. Sarbanes-Oxley (SOX) contrôle également la plupart des initiatives internes d'Autodesk.

La plate-forme Denodo permet la gestion de ces et plusieurs autres initiatives de conformité qui correspondent à d'autres lois sur la confidentialité. En utilisant la plate-forme Denodo comme un entrepôt de données logique, Autodesk a créé un seul point d'accès pour toutes les données utilisées dans l'entreprise



### AVANTAGES

Autodesk a transformé avec succès leur modèle d'affaires en utilisant la plate-forme Denodo pour la virtualisation de données. L'avantage commercial majeur de la mise en œuvre Denodo a facilité la transition d'un modèle à base de licence perpétuelle à un modèle plus moderne sur abonnement. Ce changement a amélioré la performance de l'entreprise d'Autodesk à travers l'organisation, a renforcé la collaboration en partageant l'information en temps réel dans les entreprises et l'informatique, et modernisé leur entreprise à tous les niveaux afin d'assurer une plus grande agilité, performance et rentabilité. En outre, la solution a permis à des partenaires externes d'Autodesk de se connecter directement à l'application SAP ECC d'Autodesk à travers des services Web. Cette capacité leur a permis de traiter les paiements à proximité en temps réel. Autodesk n'a pas plus besoin de déplacer des données - ils peuvent utiliser moins d'ETL. Cela implique la réduction des copies de données. En outre, pour la première fois, Autodesk est en mesure d'effectuer le point d'application unique pour la sécurité et un environnement uniforme pour l'accès aux données en place. La solution Denodo a fourni à l'équipe de développement la flexibilité nécessaire pour comprendre ce dont ils ont besoin pour construire avant de pouvoir effectivement construire. Cela a permis à l'équipe de développement de planifier à l'avance, et voir ce que leurs équipes commerciales nécessitent à partir du produit avant de commencer. Cette approche a fourni l'agilité et a rendu la virtualisation de données attrayante pour la société. La solution de virtualisation de données a également aidé les équipes d'affaires et de développement d'Autodesk à devenir plus collaboratives.





# Êtes-vous en mesure de réaliser le potentiel maximum de votre Big Data ?

*Voir pourquoi la virtualisation de données est essentielle dans votre voyage Big Data*

—  
Gouverner votre lac Big Data en utilisant la virtualisation de données

—  
Analyse de puissance avancée en utilisant les lacs de données logiques

—  
Vue 360 de votre organisation avec la virtualisation de données

—  
Fournir du contexte à votre Big Data pour mieux segmenter votre client



Denodo reconnu comme visionnaire leader dans le cadre magique de Gartner en 2016 pour les outils d'intégration de données

Denodo a été attribué le DBTA les 100 entreprises qui comptent le plus dans les données en 2016



## DIGDASH ENTERPRISE EST UN LOGICIEL DE BUSINESS INTELLIGENCE AGILE (DATAVIZ, DATA DISCOVERY ET ANALYTIQUE) CENTRÉ SUR L'UTILISATEUR MÉTIER PERMETTANT AUX ENTREPRISES DE BOOSTER LEUR PERFORMANCE.

Aujourd'hui, plus de données signifie plus de possibilités, à condition de pouvoir les traiter et les comprendre. Les rapports d'activités mensuels et les solutions orientées experts peu agiles ne suffisent plus à faire face aux très gros volumes de données (Big Data). DigDash Enterprise permet aux utilisateurs non informaticiens de faire parler leurs données instantanément et de prendre les bonnes décisions au bon moment.

DigDash a été fondé par une équipe d'experts en informatique décisionnelle. Réunis autour d'une vision plus agile de la Business Intelligence, DigDash édite DigDash Enterprise, un logiciel de DataViz permettant de créer des tableaux de bord performants et simples à utiliser.

DigDash Enterprise se connecte facilement à tous types de données structurées ou non. Grâce à ses multiples connecteurs, le logiciel fonctionne aussi bien avec des bases SQL, des cubes OLAP, des sources "big data" et des fichiers. Il complète également les solutions décisionnelles déjà en place et vous permet de créer simplement des "analytics" puissants.

Vos collaborateurs visualisent leurs données sans l'aide du service informatique et communiquent entre eux directement via les tableaux de bord. Les fonctionnalités avancées sont disponibles sans achat de modules supplémentaires.

En vous donnant accès à l'état de votre activité et à vos objectifs en temps réel, DigDash Enterprise utilise intelligemment vos données pour booster votre performance.

### TÉMOIGNAGE :

#### LA CAISSE D'ÉPARGNE DÉCUPLE SES CAPACITÉS DE PRODUCTION GRÂCE À LA BI.

Retour d'expérience de Pascal FOURNIER, Responsable département développement et décisionnel - Direction des systèmes d'information au sein de la Caisse d'Épargne Aquitaine Poitou-Charentes autour de la solution DigDash Enterprise.



### QUELS ÉTAIENT VOS BESOINS INITIAUX ?

«Pour assurer la restitution de données décisionnelles auprès de nos 2 800 utilisateurs, nous nous sommes mis en quête de la solution "idéale" qui viendrait remplacer les développements sur-mesure assurés jusqu'à présent par notre équipe. Nous souhaitons que cette solution soit une réelle plus-value pour nos utilisateurs en leur permettant d'aller encore plus loin que ce que nous leur proposons jusqu'à présent, dans l'exploration de leurs données ; et ce de manière simple, intuitive et dynamique...»

### POURQUOI LE CHOIX DIGDASH ENTERPRISE ?

«Parmi nos nombreux prérequis, nous souhaitons conserver la conception et le développement des tableaux de bord à notre niveau et pouvoir faire évoluer la solution par nos propres moyens. Après avoir vu un certain nombre de solutions, nous ne nous sommes pas contentés d'une simple démonstration de l'éditeur. Nous avons échangé par téléphone avec certains DSI équipés de DigDash Enterprise et ils nous ont tous rassuré quant à la capacité de la solution à répondre à nos besoins, ainsi que sur sa simplicité d'utilisation. De plus, la mécanique de DigDash est extrêmement puissante et de surcroît sur des serveurs raisonnables en termes de calibrage.»

«Nous venions donc de trouver la solution de Business Intelligence répondant en tous points à nos attentes en termes de fonctionnalités, de simplicité d'utilisation et de services offerts aux clients !»

### ET À CE JOUR, QUELS SONT VOS RETOURS SUR LA SOLUTION ?

« DigDash est une solution vraiment mature qui nous permet d'exploser en termes de productivité. En quelques minutes, nous faisons un tableau de bord là où il nous fallait plusieurs jours voir semaines pour le réaliser auparavant. Nos temps de développement sont quasiment divisés par 10 ! DigDash Enterprise représente pour nous le moyen de changer et d'optimiser certains modes opératoires de nos collaborateurs. En plus d'être un puissant outil de restitution, DigDash est une réelle solution d'optimisation de nos processus ! »





De la visibilité  
sur des milliards de données.



Pour en savoir plus : [www.digdash.com](http://www.digdash.com)





## HISTORIQUE

MediaSchool Group, créé en 2002 par Franck Papazian, est un groupe d'enseignement supérieur privé indépendant, spécialisé dans les domaines de la **communication**, du **journalisme**, du **digital** et de l'**audiovisuel**. Avec 21 écoles - ECS, IEJ, #SUPDEWEB et SUPDEPROD - situées dans neuf villes, et deux centres de formation continue - dont MediaSchool Executive Education dont il est question ici - MediaSchool Group forme tout au long de la vie à des métiers de passion.

## COEUR D'ACTIVITE

**MediaSchool Executive Education** est le service de **formation continue et Executive** de MediaSchool Group. Il décline les expertises thématiques des écoles - communication, journalisme, digital et audiovisuel - en formations pour les professionnels en activité.

**Le digital** requiert de nouvelles connaissances et compétences, pour pouvoir rester indispensable dans son activité. Elles sont au cœur de notre offre de formations, de la sensibilisation à l'approfondissement opérationnel et stratégique.

## PRESTATIONS / PRODUITS / SERVICES PROPOSES

Nous proposons deux **Executive Mastères (diplômants)** à temps partiel, mais aussi des séminaires courts, formations sur-mesure, cycles certifiants et des voyages d'études consacrés aux tendances digitales de demain et d'ailleurs (Etats-Unis, Chine, Canada, Israël). Les formats sont variés pour permettre à chacun de trouver son espace de respiration intellectuelle ou de travail sur des compétences opérationnelles, tout en conservant une activité intense.

## SECTEURS CIBLES

Notre thème central est **le digital**, des outils aux stratégies de communication/marketing, en passant par la conduite du changement. Il touche tous les secteurs. Historiquement implanté dans l'univers des médias, MediaSchool Group a ouvert ses secteurs cibles en Executive Education, avec des participants provenant de l'industrie, du luxe, de la santé, de la protection sociale, de l'IT, du consulting, etc.

## COUVERTURE GEOGRAPHIQUE

MediaSchool Group est implanté en France à : Paris (16<sup>e</sup>), Strasbourg, Toulouse, Marseille et Nice. Et à l'international à : Londres, Bruxelles, Barcelone, Shanghai. Les formations pour les professionnels peuvent avoir lieu soit dans ces villes, soit dans vos locaux - en France ou à l'étranger.

## NOUVEAUTES PRODUITS ET SERVICES

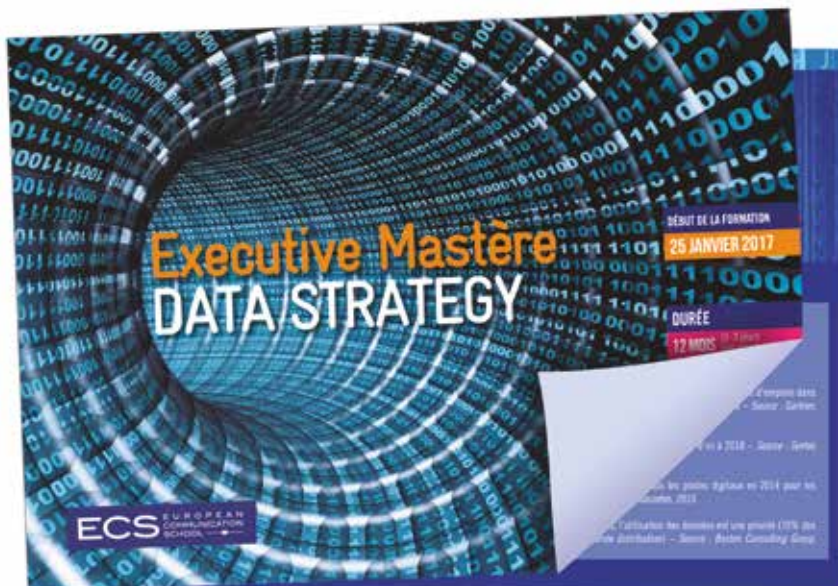
Nos deux Executive Mastères innovants s'intitulent : « **Communication, Marketing et Transformation digitale** » et « **Data Strategy** ». Sur ces Executive Mastères nous recrutons des professionnels de haut niveau, en recherche d'expertise renforcée, d'évolution, d'ouverture vers des compétences nouvelles (pour le Data Strategy) ou parfois simplement souhaitant valoriser par un diplôme reconnu de niveau I une expérience riche mais non « certifiée ». Ce type de formation représente également pour certains l'occasion de réfléchir à leur positionnement professionnel, et, tout au long du cursus, outre les échanges fertiles entre les participants et avec les intervenants, nous proposons un accompagnement personnalisé à la gestion de carrière et au Personal Branding.

Nous proposons également cette année un parcours de trois jours sur-mesure composé d'une start-up expedition, d'un Design Sprint, de séances sur le digital et d'autres séquences très innovantes.

## AVANTAGES CONCURRENTIELS

Dans le domaine du digital, MediaSchool Executive Education tire sa légitimité de **l'ancienneté de son réseau** puisqu'en formation initiale le M1 'Communication digitale' a été créé en 2009. Notre vivier d'intervenants est de ce fait extrêmement riche, composé d'**experts praticiens** qui interviennent dans nos écoles ou dans l'Institut Multi-Médias (pour cadres dirigeants) existant depuis 1983. Les formations sont évaluées et enrichies chaque année, les formateurs sélectionnés pour leur expertise, leur capacité à être en prise sur les enjeux actuels et leur pédagogie.

Un **Conseil scientifique** contribue à la définition des programmes nouveaux et à leur parfaite adéquation aux besoins des professionnels en activité. Les membres sont des professionnels de haut niveau des secteurs de la communication, des médias, du digital et des RH. Enfin MediaSchool Executive Education s'est bâti sur « l'esprit start-up », pour faire preuve de souplesse, de réactivité et de créativité.





**Julie THINÈS, créatrice et responsable de l'Executive Mastère DATA STRATEGY de l'ECS, directrice de Mediaschool Executive Education.**

**Julien MURESIANU, fondateur de JALGOS - start-up et centre de R&D en Intelligence Artificielle - et enseignant dans l'Executive Mastère DATA STRATEGY.**

#### POURQUOI SE FORMER A LA DATA STRATEGY ? (Julie THINÈS)

Aujourd'hui les données (data) permettent un fort développement de l'activité car elles peuvent être mieux et plus rapidement traitées. Pour 60% des dirigeants (selon le Boston Consulting Group, 2014), l'utilisation des données est une priorité (70% des dirigeants dans la grande distribution). Il ne s'agit donc pas de théoriser sur le 'Big Data' mais d'étudier les technologies qui permettent de les collecter, les trier, les exploiter et les protéger, afin d'accroître la performance de son organisation. Il faut également connaître les différentes catégories de data, structurées, non structurées, sociales, open data... pour pouvoir fonder une stratégie pertinente. Avec la data strategy, on parle finalement de stratégie d'entreprise. Les données en sont simplement un élément qui devient crucial car générateur de valeur et différenciant. Toutefois la plupart des entreprises ont du mal à appréhender la stratégie data car celle-ci implique un changement d'organisation, une étude fine de ROI, et surtout **d'avoir en interne des personnes compétentes pour traiter ces sujets**. Or, les compétences manquent cruellement dans le domaine. D'où notre travail sur la mise en place de cette formation Executive. Il faut rapidement former des managers capables de piloter la stratégie data avec une feuille de route claire, précise et chiffrée. Cela repose sur **une bonne connaissance de la data science, appliquée à des cas business**.

#### QUELS SONT LES ATOUTS DE L'EXECUTIVE MASTERE DATA STRATEGY ? (Julie THINÈS)

Au-delà des aspects statistiques, technologiques, stratégiques et des outils à connaître, dans un positionnement très « business/marketing », cette formation insiste aussi sur les savoir-être à mobiliser dans le cadre d'une activité liée aux data : le leadership et la collaboration - entre les différents métiers et fonctions (IT, finance, marketing...). **Nous marions la science des data avec la culture de l'agilité pour former les meilleurs managers de demain**. Ceux qui auront une « double casquette » avec une **expertise technologique et une excellente connaissance de leur activité Business**. Ils seront aussi des communicants pour expliquer leur démarche et les résultats attendus, capables de faire du data design pour fédérer les différentes ressources internes. La data strategy mobilise des compétences larges. C'est **le seul programme à ce jour qui forme des professionnels à l'ensemble de ces compétences**. Son format est également un atout : l'Executive Mastère DATA STRATEGY totalise 23 journées de formation (184 heures). Les enseignements ont lieu pendant la semaine, à raison de deux à trois jours par mois, dans nos locaux du 16<sup>e</sup> arrondissement. Un MOOC est proposé en amont pour commencer à s'immerger dans les notions de data science indispensables. Les participants pratiquent les outils, traitent des études de cas, sont mis en situation. La formation est très pratique et se conclut par la production d'une note opérationnelle.

#### A QUI S'ADRESSE CETTE FORMATION ? (Julien MURESIANU)

A des professionnels expérimentés ayant au moins cinq ans d'expérience professionnelle. Plus précisément, nous souhaitons une promotion diverse dans les métiers d'origine car les « data stratèges » doivent avant tout avoir une grande curiosité et une appétence pour les nouvelles technologies. Qu'ils viennent du marketing, de l'IT, de la stratégie ou de la production, **nous recherchons des personnes très motivées, souhaitant devenir des chefs de projet data efficaces**. Aujourd'hui selon le niveau de maturité des entreprises, la responsabilité de la stratégie data et son application incombent soit à la direction marketing soit à la DSI. Mais le chef de projet data doit avoir une vision très large. **Nous formons les interfaces entre les purs Data Scientists et les métiers**.

#### QU'ENSEIGNEZ-VOUS DANS L'EXECUTIVE MASTERE « DATA STRATEGY » ? (Julien MURESIANU)

J'enseigne les grands principes de l'analyse statistique (statistiques traditionnelles, modèles de base et utilisation en entreprise), la Business Intelligence, le Data-mining et le Model-thinking. J'explique le plus simplement possible les familles d'algorithmes et leur utilisation : optimisation, prédiction, clustering, matching et design d'algorithmes. Mes enseignements sont conçus pour être accessibles à des non-scientifiques. J'utilise beaucoup d'images et d'exemples pour faire comprendre la data science à ces cadres qui viennent de différents secteurs d'activité. J'anime également une séquence sur l'éthique car les algorithmes posent un certain nombre de dilemmes moraux. L'idée n'est pas de donner une réponse toute faite mais de sensibiliser les participants au réflexe de se poser des questions en la matière afin qu'ils puissent décider en toute conscience de l'utilisation qu'ils vont faire des données dans leurs futurs postes.

#### QUE PENSEZ-VOUS DES PARTICIPANTS A CE PROGRAMME ET COMMENT LES VOYEZ-VOUS EVOLUER ? (Julien MURESIANU)

Je suis très agréablement surpris par le niveau de la promotion sélectionnée cette année. Ils sont hyper investis, très dynamiques, passionnés par le sujet. Ils font des choses, posent beaucoup de questions, veulent comprendre, c'est vraiment très plaisant. L'ambiance est excellente et j'ai vraiment senti entre mes deux interventions le progrès qui a été fait. Ils ont acquis des réflexes et sont bien partis pour être à la sortie d'excellents Chief Data Officers ou équivalents.

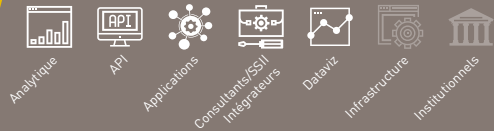
**François POITRINE**

Directeur Général - Responsable de la pratique Big-Data

+33(0)1 70 82 52 71

fp@ekimetrics.com

136 - avenue des champs Elysées 75008 Paris France



www.ekimetrics.com



# Ekimetrics.

## HISTORIQUE

Ekimetrics est un cabinet de conseil en stratégie spécialisé en Data Science. Le cabinet a été fondé en 2006 à Paris et compte à présent plus de 150 consultants basés à Paris, Londres, New York, Hong Kong et Dubaï. L'approche pionnière et rigoureuse dans le traitement de la data a permis à Ekimetrics de devenir le leader européen indépendant du ROI et du conseil en Data Science.

## COEUR D'ACTIVITE

Ekimetrics couvre l'ensemble de la chaîne de valeur de la data et déploie une expertise avancée dans le conseil en technologie, la gouvernance des données, le cadrage stratégique et l'utilisation des algorithmes de machine learning à des fins business et marketing. Les consultants d'Ekimetrics disposent d'une double expertise business et mathématiques appliquées, pour identifier et mettre en place les méthodologies les plus adaptées aux questions posées.

## PRESTATIONS / PRODUITS / SERVICES PROPOSES

Chaque modèle créé adresse une question spécifique soulevée par les clients et tient compte des particularités business en découlant : marché, chaîne de valeur, accessibilité des données, etc. La mesure de performance est pensée en amont de chaque projet. Ekimetrics privilégie une approche agile, avec des boucles de modélisation et de test & learn.

## SECTEURS CIBLES

Une approche agile et des méthodologies propriétaires permettent à Ekimetrics de s'adapter à une grande variété de problématiques et d'industries. Le cabinet conseille depuis plusieurs années des leaders européens et internationaux issus de secteurs aussi variés que le luxe, l'automobile, la banque-assurance, la cosmétique, la grande consommation ou l'énergie.

## COUVERTURE GEOGRAPHIQUE

Historiquement Ekimetrics est très présent en Europe, le cabinet a su évoluer et couvrir différentes zones pour accompagner ses clients dans leur stratégie internationale. A présent doté de cinq bureaux : Paris, Londres, New York, Hong Kong et Dubaï ; Ekimetrics bénéficie d'une réelle couverture à l'international et compte parmi ses salariés plus de 20 nationalités.

## F. NOUVEAUTES PRODUITS ET SERVICES

L'équipe Data Operation Team qui s'occupe des projets BI & Big Data au sein d'Ekimetrics se renforce continuellement pour assurer la réalisation et le déploiement de plusieurs DAP - Data Analytics Platform - et DMP - Data Management Platform - pour différents clients. Ekimetrics est actuellement à la pointe dans l'utilisation des distributions Hadoop ainsi que Spark. Cette année Ekimetrics a aussi étendu ses équipes en Machine Learning et en product engineering pour répondre aux besoins croissants de ses clients dans les domaines algorithmiques et applicatifs - applications web, dataviz, outils d'aide à la décision. La capacité à concevoir et déployer des architectures Big Data complexes conjuguée à la maîtrise du machine learning et des langages applicatifs permet à Ekimetrics d'offrir une expertise unique en Data Science dans l'hexagone.

## AVANTAGES CONCURRENTIELS

Ekimetrics est indépendant de tout organisme de médias et d'éditeurs de logiciels, ce qui permet au cabinet une grande agilité dans le choix des outils et des méthodes, ainsi qu'une réelle impartialité dans les recommandations. Les 150 consultants ont tous un double profil consultant / statisticien-Data Scientist (X, ENSAE, Telecom, Mines, Columbia, Imperial College). Ekimetrics centre son approche sur les besoins business des clients pour construire la méthodologie et l'architecture technique adaptée. Grâce à une triple expertise stratégique, statistique et technologique, les consultants sont à même de livrer des recommandations actionnables par les CIOs, CDOs et les CMOs. Ekimetrics accompagne ses clients pour optimiser l'ensemble de la chaîne de valeur de la data et joue un rôle d'interface entre les métiers, la DSI et les analytics sur des projets data transversaux.

## Big Data Expertise

Machine Learning  
Text Mining  
Data Viz



## Conseil technologique



Faciliter le choix des composantes technologiques en fonction du besoin. Conseil en architecture Big Data.

## Exploitation de données



Développer en environnement agile et peu coûteux pour réaliser des PoCs. Industrialisation des solutions performantes

## Cadrage stratégique



Choisir les questions business les plus critiques à adresser. Cadrage fonctionnel et budgétaire.

## Pilotage du Marketing Mix par la performance



Media



Pricing



Digital



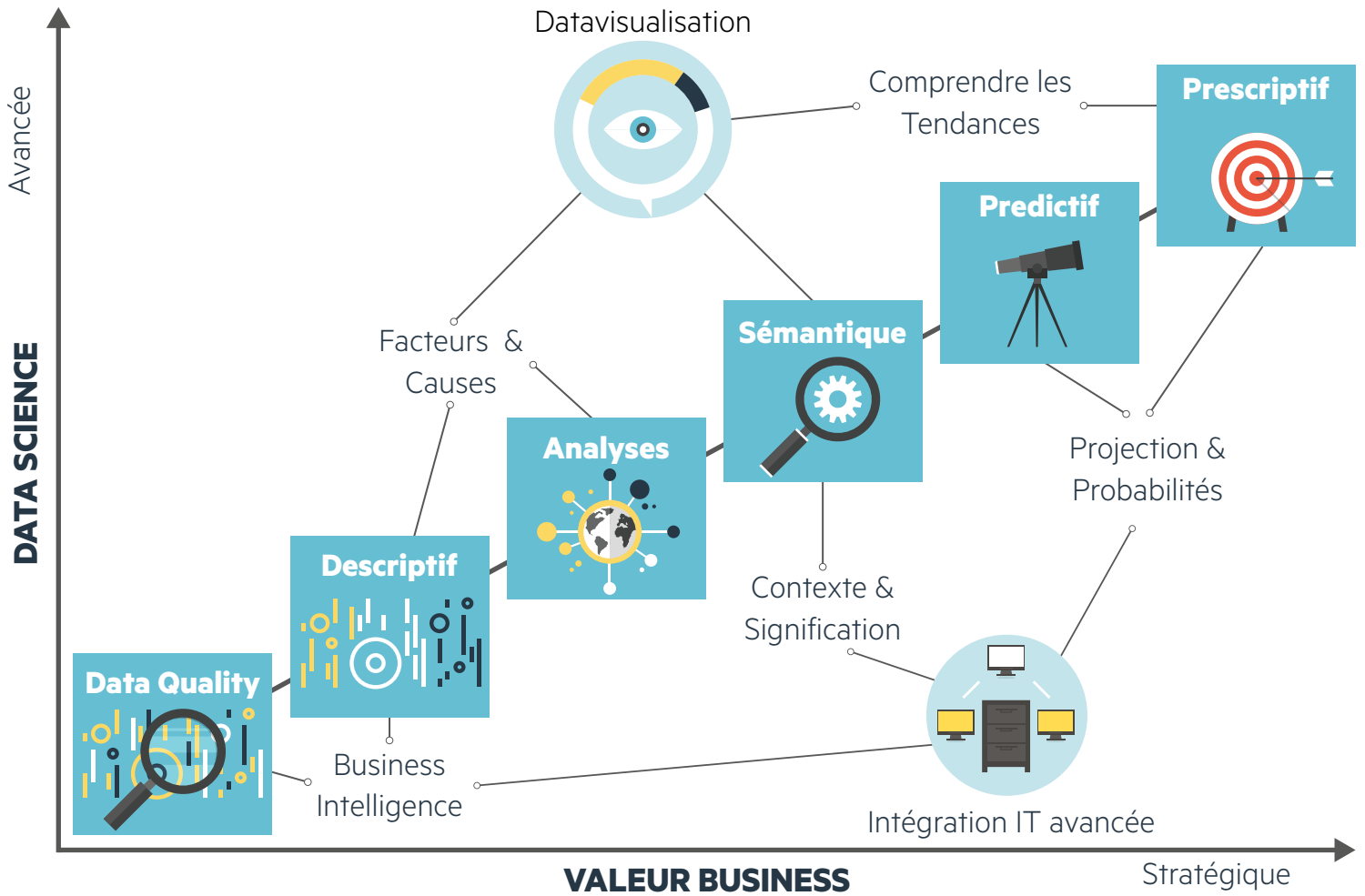
Trade Marketing



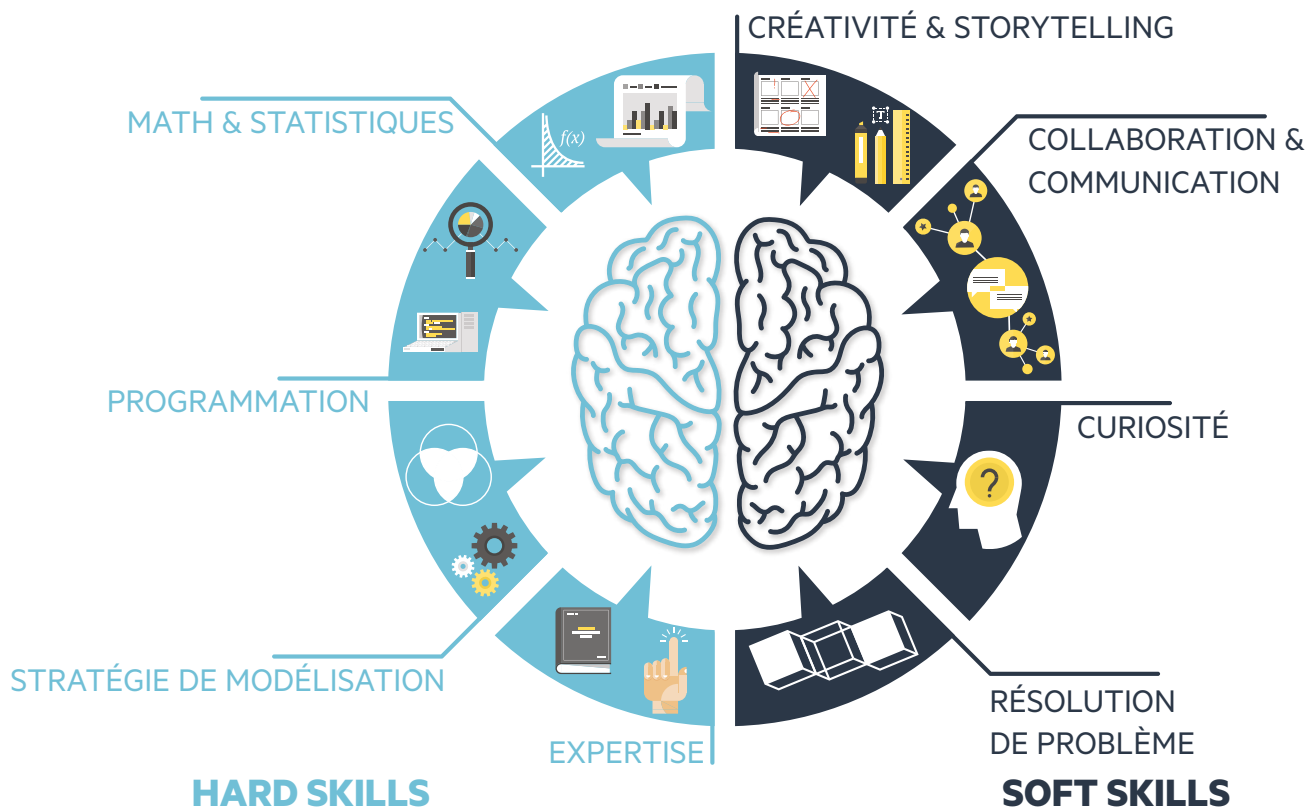
CRM

# Expertise Big Data Ekimetrics.

UNE MAITRISE TRANSVERSALE DE LA CHAÎNE DE VALEUR DE LA DATA



## Dans le cerveau d'un Data Scientist





elastic

**L'ÉVOLUTION DE LA FONCTION DE RECHERCHE, AU-DELÀ DE GOOGLE**

Si l'on repense aux méthodes de recherche, les consommateurs et entreprises d'il y a quelques années étaient contraints de consulter les données d'une manière précise. Dans l'annuaire papier des Pages Jaunes, par exemple, la recherche d'une entreprise se faisait par ordre alphabétique, et il était impossible d'enrichir sa recherche avec d'autres critères, tels que la taille, le chiffre d'affaires, l'influence ou d'autres champs pertinents.

Plus récemment, Google est l'outil de recherche auquel nous nous référons le plus souvent : pour y saisir un mot ou une suite de mots et obtenir en retour une liste de résultats correspondants. Une fois les résultats affichés, l'utilisateur fait défiler la page jusqu'à trouver ce qu'il recherche. Même si cet outil est un moteur de recherche bien plus efficace que celui de l'annuaire papier des Pages Jaunes, il contraint néanmoins l'utilisateur à une consultation des données assez similaire, par un affichage « unique ».

Elasticsearch élimine cette restriction en affichant un groupe de données synthétisées, multidimensionnelles et issues d'un éventail de sources variées, le tout en temps réel. L'utilisateur est alors en mesure de déceler les informations correspondant à ses besoins, à partir d'une quantité quasi infinie de données. Cette solution permet également aux utilisateurs de réaliser des analyses et des graphiques de visualisation des données. L'affichage des informations est alors plus accessible qu'une liste de résultats et permet d'exécuter des analyses et des rapports à une vitesse éclair. De nombreuses applications qui rythment notre quotidien s'appuient déjà sur cette solution, notamment BlaBlaCar, Uber, Facebook, Netflix, Wikipédia et des systèmes de trading de grande envergure chez Goldman Sachs ainsi que d'autres institutions financières telles que la Société Générale ou la Banque de France..

**LA SUITE ELASTIC : UNE SOLUTION QUI S'ADAPTE À TOUS LES CAS D'UTILISATION ET À TOUT LE MONDE**

La Suite Elastic (Elasticsearch, Kibana, Beats et Logstash) constitue une approche ouverte et hautement distribuée qui est de mise chez des entreprises aux besoins variés et dotées d'infrastructures complexes, pour des cas d'utilisation très différents, que ce soit la recherche, l'analyse, le logging ou encore sécurité. Qu'elle soit déployée sur site ou dans le cloud, la Suite Elastic s'avère être l'outil le plus performant pour des entreprises de toutes sortes, leur permettant de rassembler, d'analyser et d'extraire une valeur en temps réel d'ensembles de données volumineux et souvent hétérogènes.

**OFFRIR À L'ENTREPRISE DE LA VALEUR IMMÉDIATE**

De nos jours, la masse de données disponible signifie qu'il est d'autant plus vital pour les développeurs de créer des applications capables d'offrir une valeur immédiate à leurs entreprises. Grâce à la Suite Elastic, aux fonctionnalités X-Pack (security, monitoring, alerting, reporting et Graph) les développeurs et utilisateurs finaux sont en mesure d'injecter de la valeur quel que soit leur cas d'utilisation. Voir les exemples ci-dessous :

Recherche intégrée	Analyse de logs	Analyse sécuritaire
<ul style="list-style-type: none"> <li>• Personnaliser l'expérience de recherche afin de booster l'adoption par l'utilisateur</li> <li>• Faciliter des stratégies de monétisation en temps réel, que ce soit par la publicité ou les achats « in-app »</li> <li>• Connaître les actions des utilisateurs pour offrir le meilleur service client.</li> </ul>	<ul style="list-style-type: none"> <li>• Optimiser la performance système, diagnostiquer les problèmes en temps réel et gérer l'évolutivité</li> <li>• Respecter les SLA de service client, garantir la disponibilité MTTI/MTTR, les opérations de développement à distance</li> <li>• Exploiter les données et créer de nouvelles opportunités de revenus à l'aide d'indicateurs.</li> </ul>	<ul style="list-style-type: none"> <li>• Déceler automatiquement les anomalies et les tendances cachées au sein des données susceptibles de poser un risque sécuritaire</li> <li>• Permettre la détection des menaces en temps réel pour les initiatives de sécurité informatique et de lutte contre la fraude</li> <li>• Respecter les normes de conformité, limiter les risques financiers et les risques utilisateurs.</li> </ul>

Pour découvrir comment Elastic peut vous aider et en savoir plus à propos de nos offres de souscription, n'hésitez pas à nous contacter.

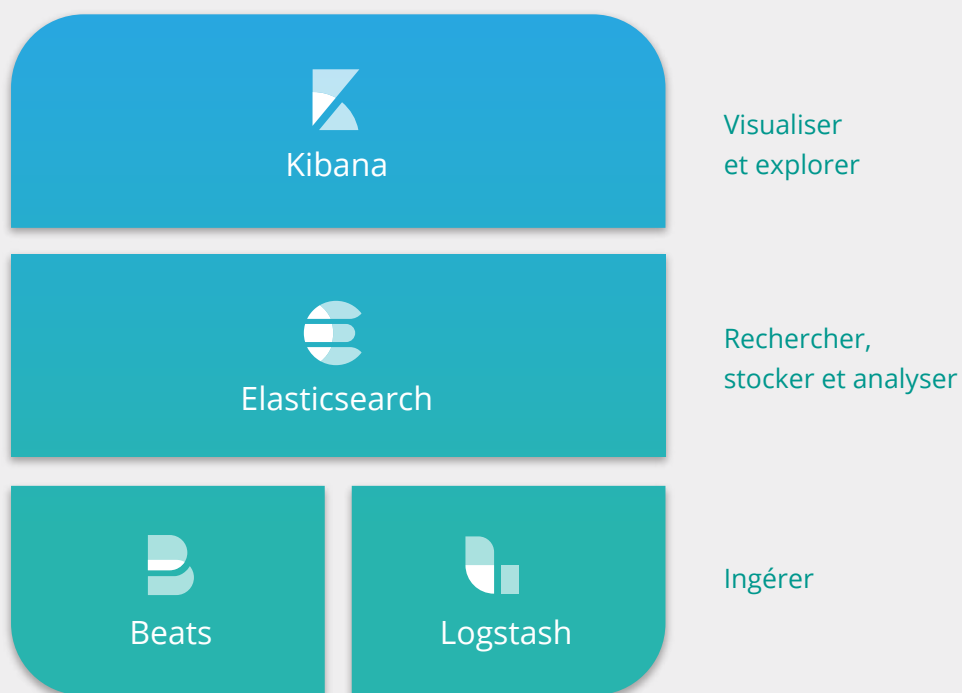
Ou testez Elastic Cloud gratuitement pendant 14 jours pour découvrir le potentiel de la Suite Elastic par vous-même : <https://www.elastic.co/cloud/>





# LA SUITE ELASTIC

Open source, flexible et super rapide.  
Utilisez la sur vos infrastructures ou dans le cloud.



Pour trouver une solution à vos problèmes opérationnels, il faut commencer par chercher. La Suite Elastic est la solution complète qui fera de vos données un atout exploitable en temps réel pour des cas d'utilisation stratégiques, que ce soit la recherche, l'analyse, le logging ou encore la sécurité.

**Essayez Elastic Cloud  
gratuitement pendant 14 jours**  
et explorez le potentiel de la Suite  
Elastic par vous-même.  
[www.elastic.co/cloud](http://www.elastic.co/cloud)

**Contactez-nous pour plus de  
détails concernant nos offres  
de souscription.**  
+33 1 76 75 33 03  
[www.elastic.co/fr/](http://www.elastic.co/fr/)





## FRAUD ANALYTICS – RISK MANAGEMENT

« Eulidia nous a aidé à lutter contre la fraude en transformant nos pratiques par le levier de la Data Science, guidé par le management de la performance »

**La démarque inconnue grève de plus de 1% le chiffre d'affaire des distributeurs...**

Une des missions de l'audit interne est de lutter efficacement contre ce phénomène. Une fois un schéma de fraude connu, cette équipe analyse et cible ses audits en croisant les données de stock et d'encaissement accessibles via des technologies classiques de Business Intelligence. Pour traiter les nouveaux cas de fraudes, les auditeurs s'appuient sur la DSI pour toutes nouvelles études ad hoc. Avec l'augmentation des volumes et le besoin d'être bien plus réactif face à l'intelligence des fraudeurs, l'audit a exprimé son besoin d'augmenter sa capacité de détection et d'analyse.

**Passer d'un modèle réactif à un modèle proactif**

Pour augmenter l'autonomie du métier et faire évoluer ses pratiques de prévention et d'amélioration continue (benchmark & pilotage), nous avons imaginé un moyen innovant de détecter les anomalies à partir des données de manière proactive et instauré de nouveaux usages des technologies BI & Big Data.

**Une démarche d'innovation rapide**

Nous avons opté pour la mise en œuvre d'un « Proof of Value industrialisable » sur un périmètre restreint. Après un cadrage du business case, une démarche sur mesure a été validée avec le client. Cette démarche s'est appuyée sur notre offre Risk Management et plusieurs de nos méthodes. Enfin, en collaboration avec la DSI, nous avons mis en œuvre et hébergé une infrastructure Big Data dédiée pour lancer rapidement le POV et lever toutes contraintes techniques.

**Un travail en équipe pluridisciplinaire**

Notre équipe pluridisciplinaire a conjugué les compétences d'un architecte Big Data, d'un Data Scientist senior, d'un Partner conseil en Management. Elle a pu implémenter en quelques jours le socle Big Data sur Cloudera / Impala / Hive et l'alimenter de données utiles. Le Data Scientist a utilisé R pour les analyses et la modélisation, et enfin Tableau Software pour la BI self-service & la Data Visualisation. L'équipe a travaillé en proximité avec le responsable de l'audit et le responsable BI pour garantir la bonne compréhension métier, livrer les insights, assurer le bon déroulement du projet et la conduite du changement.

**La transformation de la DATA en insights activables**

La démarche a permis, de mieux comprendre le phénomène de la fraude (Insights & Dataviz), de mettre en place les moyens de détecter de nouveaux risques (Advanced Analytics), et enfin de proposer un plan d'actions pertinent et son dispositif de mise sous contrôle (Business Performance).

**Notre création de valeur**

Nous avons prouvé en l'espace de 2 mois, que le métier peut :

- Passer d'une approche réactive à une approche proactive du management des risques grâce à la détection de comportements atypiques à partir de signaux faibles,
- Analyser et comprendre en autonomie les tendances et la répartition des opérations à risques (cartographie des risques interactive) afin de mieux cibler les zones prioritaires à auditer (gains de temps & coûts de déplacement),
- Conduire des benchmarks entre magasins visant à harmoniser les pratiques en caisse,
- La DSI dispose désormais d'une infrastructure de POV Big Analytics.

## PROCESS ANALYTICS – PERFORMANCE DIGITALE

« Eulidia nous a aidé à augmenter le taux de transformation du processus d'acquisition client par le levier du Process Mining »

**Maximiser le R.O.I. des investissements consacrés au lead generation...**

Notre client, une entreprise digitale, souhaite augmenter le taux de transformation de son processus d'acquisition allant du lead web à l'ouverture d'un service. Ce processus connaît des difficultés sur sa performance. Alors que l'approche classique lean 6 sigma atteint sa limite, le défi est de trouver les leviers d'actions qui permettent d'augmenter significativement son taux de transformation.

**Dépasser les limites du Web Analytics**

Notre démarche unique s'est appuyée sur les pratiques du Process Mining, du Lean Management et de la Performance Marketing. Elle innove en permettant à nos clients d'explorer, d'améliorer et de piloter le processus réel par l'utilisation des logs des différents canaux. Pour l'exploration des parcours, plusieurs algorithmes permettent de créer un modèle de graphe à partir des schémas les plus fréquents. La combinaison des pratiques de Performance Management et des techniques Data Science a permis d'avoir une vision compréhensible des parcours et des interactions réellement mis en œuvre avec le client, de les optimiser pour mieux répondre aux attentes et enfin de les piloter.

**Une démarche d'innovation rapide**

Cette démarche suit le cycle DMAIC (Define, Measure, Analyse, Improve, Control). Le cycle Define a embarqué le cadrage et la réalisation d'un Proof of Value. Une fois cadré nous avons mis en place le socle de collecte, la visualisation du parcours tel qu'il existe sans a priori. Les analyses ont été faites à l'aide de Disco et R et le pilotage de la performance du parcours à l'aide de Tableau Software.

**Un travail en équipe pluridisciplinaire**

Notre équipe pluridisciplinaire (BI, Data Science et Conseil en Management) a accompagné le marketing et l'excellence opérationnelle pour conduire l'analyse exploratoire du parcours, identifier les familles de sous-processus par clustering, évaluer leur performance, comprendre les points de blocage, détecter les profils à risque et proposer un plan d'actions associé à chaque opportunité d'augmentation de transformation.

**Notre création de valeur**

Nous avons créé de la valeur par :

- Une lecture simple sur la variété des parcours réels,
- Une connaissance des variables explicatives qui impactent le taux de transformation,
- L'identification des points faibles et zones à risques des parcours,
- Des leviers de transformation en fonction des opportunités de gains, et des parcours cibles optimisés en fonction de la segmentation client.





**WE ARE**

**BUSINESS PERFORMANCE PARTNERS**



# eulidia

BUSINESS PERFORMANCE PARTNER

TURN **DATA** INTO  
**BUSINESS**  
**VALUE**

Mixing Business Intelligence & Data Science

**INNOVATION** | **DATA FACTORY** | **ADVANCED ANALYTICS** | **BUSINESS INSIGHTS** | **DATAVIZ**

[www.eulidia.com](http://www.eulidia.com)

## Axel BRANGER

Sales Director EXALEAD - South Europe - Russia & Middle East

+33 (0) 1 55 35 26 26

axel.branger@3ds.com

10 place de la Madeleine 75008 Paris France



[www.3ds.com/fr/produits-et-services/exalead](http://www.3ds.com/fr/produits-et-services/exalead)



## HISTORIQUE

Dassault Systèmes est un éditeur de logiciel français né en 1981 et aujourd'hui Société Européenne.

EXALEAD a été créée en 2000 par deux pionniers des moteurs de recherche. Rachetée en 2010 par Dassault Systèmes, EXALEAD est la marque dédiée à la création des solutions d'exploration et d'analyse de l'information pour mettre les données au cœur de la transformation des entreprises.

## COEUR D'ACTIVITE

EXALEAD Dassault Systèmes offre une technologie de recherche et d'indexation, couplée à des technologies web, pour un accès universel et unifié et un traitement haute-performance de l'information.

Ses solutions permettent :

- d'accéder agilement et rapidement à un grand volume de données dispersées et hétérogènes
- d'exploiter cet existant pour réutiliser et analyser l'information
- de révéler les liens et le contexte des données pour créer des applications d'entreprise innovantes
- de donner du sens à l'invisible pour créer plus de valeur et répondre aux exigences de transformation

## PRESTATIONS / PRODUITS / SERVICES PROPOSES

EXALEAD adresse l'essentiel des problématiques Big Data des organisations industrielles et de services, à savoir :

PLM Search & Analytics

- Sourcing and Standardization Intelligence
- Product Development Analytics

Customer Support & Service Analytics

- Customer Engagement
- Customer Support

## SECTEURS CIBLES

EXALEAD offre des applications personnalisées ou packagées, évolutives, parfaitement adaptées aux environnements actuels de Big Data et qui permettent de le transformer en vrais bénéfices pour l'entreprise, quelle que soit sa taille.

Elles s'adressent à tous les secteurs d'activités (Banques et assurances, Telecom, Défense, High Tech, Automobile, Aéronautique, Energie, Santé, Grande distribution...) et toutes les fonctions de l'entreprise.

## COUVERTURE GEOGRAPHIQUE

Dassault Systèmes est présent partout dans le monde. Les solutions EXALEAD sont aujourd'hui déployées en Europe, Amérique du Nord, Moyen-Orient, Afrique et Asie.

## NOUVEAUTES PRODUITS ET SERVICES

Les solutions OnePart et PLM Analytics sont désormais également disponibles dans la 3DEXPERIENCE platform de Dassault Systèmes.

La nouvelle version d'EXALEAD OneCall, solution dédiée à l'interaction client, offre maintenant le module Advanced Customer Analytics qui permet d'apprendre des données pour anticiper les besoins clients, détecter leurs comportements et proposer de nouvelles offres ou empêcher l'attrition ; autant d'actions qu'il est maintenant possible de générer industriellement grâce au module analytique.

## AVANTAGES CONCURRENTIELS

EXALEAD propose un robot d'indexation haute performance, des fonctionnalités sémantiques avancées et un index Web exploitable, ainsi qu'un puissant portefeuille de connecteurs pour les sources de big data structurées et non structurées à l'intérieur et à l'extérieur de l'entreprise.

EXALEAD est une technologie évolutive, multi-sources, rapidement mise en œuvre et au TCO faible.

Les clients EXALEAD bénéficient de toute la structure et l'implantation des équipes conseil, support et formation mais également du réseau de partenaires Dassault Systèmes.

# L'intelligence de la donnée au cœur de votre transformation

- Explorer agilement et rapidement
- Exploiter l'existant
- Révéler les liens entre les données
- Donner du sens à l'invisible



**Johan CHEFDEVILLE**

BI & Big Data Practice Leader

+33(0)1 77 51 03 00

johan.chefdeville@experis-it.fr

Immeuble Eureka

13 rue Ernest Renan 92723 Nanterre Cedex - France



www.experis-it.fr



**Experis™ IT**  
ManpowerGroup

## EXPERIS IT ACCOMPAGNE SES CLIENTS DANS LEUR TRANSFORMATION DIGITALE EN RÉPONDANT À LEURS BESOINS DE COMPÉTENCES ET DE CONSEIL, SELON LES MODALITÉS LES PLUS APPROPRIÉES À LEUR CONTEXTE.

Notre Data Lab, véritable accélérateur d'incubations de projets, s'appuie sur les principes de modélisation et de prédiction pour proposer une approche métiers innovante. Il met en perspective les relations entre les données et la façon dont elles vont répondre aux enjeux Business.

Notre dernière offre, BIG DATA RH, apporte de nouveaux éclairages pour optimiser la gestion des Ressources Humaines. Grâce aux technologies du Big Data, nous proposons une solution de Matching et de Data Mining visant à aider les fonctions RH sur l'ensemble de leurs activités : recrutement, développement des compétences, gestion de carrières ou encore rémunération.

ManpowerGroup accompagne les entreprises dans la définition et la mise en œuvre de leur stratégie RH par le recours exclusif ou combiné de solutions en matière de gestion de la flexibilité, recrutement, formation, gestion des carrières, conseil en stratégie RH, gestion externalisée de fonctions RH, etc... Pour ses clients mais aussi en interne, il est apparu crucial d'apporter des solutions innovantes aux modes de gestion des Ressources Humaines afin d'optimiser les processus et générer de la valeur. Face à une fonction Ressources Humaines en pleine mutation, parent pauvre en matière de Data Science, il est fondamental d'en revoir la perception et de l'appréhender de façon plus globale. Comment faire évoluer son organisation et ses ressources ?

Notre Data Lab propose une méthodologie de screening des processus et des données adaptée aux spécificités des Ressources Humaines. L'entreprise accède à un outil d'aide à la décision permettant d'améliorer l'ensemble des processus métiers de la Gestion des Ressources Humaines. Ainsi, les opérationnels RH disposent de toutes les clés pour identifier les meilleurs candidats internes et externes par poste à pourvoir, proposer la meilleure adéquation de missions et de parcours aux salariés, détecter les hauts potentiels, ou encore les profils à fort risque de départ...

D'un point de vue purement technique, le modèle repose sur un moteur d'analyse qui permet de traiter l'ensemble des flux de données inhérentes aux collaborateurs / candidats à des fins prédictives, qu'il s'agisse de données internes à l'entreprise, ou externes, et ce quel que soit leur format : données de CV, descriptifs de postes, résultats de tests techniques, évaluations recruteurs ou RH, synthèses d'entretiens annuels, descriptifs de postes, formations passées ou planifiées, rémunération...

### Les grands principes

- Les données inhérentes aux collaborateurs / candidats sont indexées grâce à un traitement sémantique
- Les profils des collaborateurs / candidats ainsi définis sont qualifiés et évalués
- Le moteur d'analyse fait appel à un système exploratoire intelligent sur le principe du *Knowledge Discovery*. En clair, les critères de qualification sont incrémentés en continu, en tenant compte de l'émergence de nouveaux concepts (tendances, mots clés....) au fur et à mesure de l'enrichissement de la base de données.

Notre approche est résolument orientée « co-working », au plus proche des spécificités de nos clients, de manière à proposer des solutions personnalisées et évolutives, répondant efficacement aux besoins et attentes. Pour que l'intégration de l'analytique RH soit un mouvement gagnant-gagnant pour l'entreprise et les collaborateurs / candidats, il est bien évidemment nécessaire de veiller à respecter les règles du droit du travail et la réglementation française & européenne sur les données à caractère personnel. Direction Juridique et « Chief Digital Officer / Data Privacy Officer », sont toujours associés à nos projets pour assurer la régularité et la bonne gouvernance.

### Quelques exemples de scorings opérationnels pour la prise de décision, avec des résultats en 1 CLIC !

- Degré de séniorité : évaluation de la richesse de l'expérience
- Tops profils : identification de parcours exemplaires
- Indice de rétention : analyse prédictive du risque de démission

## L'incubateur Experis IT

<b>Objectif</b>	Fédérer les forces vives
<b>Description du dispositif</b>	<p><b>Une équipe</b></p> <ul style="list-style-type: none"> <li>• Noyau dur : experts métiers et data science</li> <li>• Apprentis : data analystes</li> <li>• Guests, selon projet : référents, start-ups...</li> </ul> <p><b>Des outils</b></p> <ul style="list-style-type: none"> <li>• DataViz</li> <li>• Plateforme d'exploration</li> <li>• Espaces collaboratifs</li> </ul> <p><b>Des modèles de fonctionnement</b></p> <ul style="list-style-type: none"> <li>• Méthodes agiles : itération et collaboration</li> <li>• Méthodologies d'accélération : CK, Heuristique</li> </ul>
<b>Premières étapes</b>	<ul style="list-style-type: none"> <li>• Constitution d'une équipe projet</li> <li>• Conduite des 1<sup>ers</sup> cas d'usage (POC/POV)</li> <li>• Bilan en fin de période d'exploration</li> </ul>

PROOF OF CONCEPT POC /  
PROOF OF VALUE POV

Mise en place d'un incubateur

- Intégration des données
- Analyse des données
- Visualisation et reporting



**BI • BIG DATA  
DATA SCIENCE**

# Révélez la richesse de vos données

**Nos experts interviennent pour valoriser  
l'ensemble des données de votre entreprise  
et optimiser sa performance**

Diagnostic	Data Visualisation
Gestion de Projet	Reporting
Prototypage POC / POV	Data Management

**Contactez-nous**  
**experis-it.fr • 01 77 51 03 00**

**EXPERIS IT,  
ENTREPRISES DE SERVICES DU NUMÉRIQUE  
DE MANPOWERGROUP**

- MOA, Gouvernance, Gestion de Projet
- Solutions Applicatives
- BI, Big Data, Data Science



**Experis™ IT**  
ManpowerGroup



## « C'EST D'UNE PASSION COMMUNE POUR LES MARCHÉS FINANCIERS ET L'INFORMATIQUE QU'EST NÉE FINAXYS EN 2008.

Finaxys est un cabinet de conseil en stratégie IT pour les acteurs de la finance de marché. Nos experts, recrutés pour leur appétence pour les métiers de la finance, accompagnent nos clients dans leur conduite du changement en apportant des conseils stratégiques et en implémentant des technologies innovantes. »

### Finaxys en chiffres :

2008 année de création, 31m € de chiffre d'affaires en 2015 (+20%)  
3 implantations : Paris, Londres & Bruxelles / 300 collaborateurs  
60% de projets à engagement de résultats.

### POUVEZ-VOUS NOUS PRÉSENTER FINAXYS ?

Finaxys accompagne les acteurs du secteur financier autour de problématiques de Conseil en SI en s'appuyant sur sa connaissance métier et sur une démarche qualité intégrée. A cela, s'ajoute le positionnement stratégique de nos offres qui accompagnent les enjeux de transformation de nos clients grâce à du conseil ultra-personnalisé et la mise en œuvre de solutions sur mesure. Cette combinaison d'expertises sur les technologies Big Data et le savoir-faire dans le domaine de la Finance, nous permet d'offrir à nos clients des réponses et solutions performantes parfaitement adaptées à leur environnement. Notre ADN combine agilité et innovation. Aujourd'hui, grâce à notre maîtrise des technologies Big Data et Cloud, nous restons le meilleur partenaire de nos clients sur les marchés financiers.

### POUVEZ-VOUS NOUS PARLER DE VOTRE PHILOSOPHIE AUTOUR DE LA CULTURE DATA CHEZ FINAXYS ?

Nous sommes convaincus que les solutions Big Data apportent une réponse pertinente aux problématiques IT et métiers dans le domaine bancaire, et c'est la raison pour laquelle nous nous mobilisons fortement depuis 4 ans pour leur adoption. Au-delà de la nécessaire maîtrise des architectures et des enjeux de notre secteur d'activité, nous avons une équipe R&D qui intervient au cœur des technologies, des clusters « bacs à sable » sont mis à disposition des consultants pour évaluer les technologies en toute sécurité. Nous avons développé une plateforme qui permet de mettre en œuvre les derniers composants sur des cas réels. Outre les multiples partenariats que nous entretenons activement depuis quelques années (HortonWorks, Cloudera, MapR, AWS, Couchbase, Confluent, Elastic, etc), nous sommes des intervenants convaincus, passionnés et dynamiques au sein de la communauté Big Data (Meetup Paris Hadoop User Group, Paris Datageeks, ElasticSearch, conférence Distributed Matters, Strata, etc). Ce n'est qu'ensuite que vient s'ajouter notre expertise en matière de gouvernance, qui fait de FINAXYS un acteur unique, dont la spontanéité et l'agilité structurée, remettent le principe d'écoute des interlocuteurs au cœur de nos missions de conseil. Ce principe de respect du capital humain est également au centre de notre mode de fonctionnement interne ; nous sommes des passionnés, nous aimons transmettre et nous considérons que la connaissance doit être partagée pour alimenter les expertises de l'ensemble de nos collaborateurs.

### QU'EN EST-IL DE VOTRE APPROCHE DE LA CULTURE DATA ?

La gouvernance des données est devenu un enjeu majeur avec la démocratisation des technologies permettant la manipulation de très grands volumes. Les entreprises cherchent à démultiplier les sources d'informations afin d'enrichir celles qu'elles disposent en interne. La connaissance client est au cœur des nouveaux business model bancaire. Notre volonté est de servir d'accélérateur de cette transformation en mettant la donnée

aux cœurs des actifs de la banque et en développant l'intelligence collective de l'entreprise. Notre approche ne se limite donc pas à maîtriser les technologies. Nous sommes convaincus que la création de valeur passe par trois éléments fondamentaux :

1. Mettre les données au centre des actifs de l'entreprise. Cela est rendu possible par l'implémentation d'une plateforme collaborative (Data Lake) agile et la mise en œuvre d'une architecture « Data Centric ».
2. Le succès d'une gouvernance de données agile est également lié à l'appropriation du nouveau paradigme par les collaborateurs. L'enjeu est de valoriser le partage de données par l'adoption d'une organisation « Data Driven ».
3. Toujours dans cette démarche d'agilité, nous sommes convaincus que le « Cloud » répond à la problématique de maîtrise des coûts tout en respectant les contraintes de sécurité et de réglementation du secteur.

### POUVEZ-VOUS NOUS PARLER D'UN EXEMPLE DE PROJET CHEZ UN DE VOS CLIENTS ?

L'exemple qui me semble le plus pertinent à vous raconter est celui d'une banque d'investissement qui souhaitait être accompagnée dans la construction d'un repository risque afin de répondre aux attentes des nouvelles réglementations (BCBS 239 et FRTB). En effet, l'architecture SI de calcul de la XVA de notre client n'était pas en mesure de gérer les volumes de données qui explosaient du fait de la réglementation future (FRTB) et n'avait pas non plus la flexibilité requise pour faire face à ces évolutions. La mise à niveau du système existant était impossible, il a donc fallu développer une application maison Big Data en se basant sur des composants Hadoop. A cela s'ajoutait le souhait de notre client d'utiliser l'application pour optimiser ses fonds propres à des fins réglementaires, ce qui a nécessité d'implémenter des fonctionnalités de simulation dans l'application finale. Finalement, les composants Hadoop, Spark et Kafka ont été utilisés et seront bientôt la base de la nouvelle application de gestion des risques de marché du client. En matière de scalabilité, un nombre illimité de vecteurs de sensibilités, d'agrégats et de scénarios sont gérés par la plateforme, et en terme de flexibilité, les indicateurs de risque peuvent être ajoutés ou modifiés sans coût ni interruption de service. L'architecture événementielle permet de pouvoir alimenter au « fil de l'eau » le repository tout au long de la journée et de mettre les analyses à disposition du métier au plus tôt.



Pierre Bittner, CTO FINAXYS et Responsable de l'offre de Conseil Big Data FINAXYS.



# SOMETHING BIG HAS HAPPENED !

LE BIG DATA BY



**15** CONSULTANTS  
DANS LA  
COMMUNAUTÉ  
BIG DATA  
<https://finaxys.slack.com/>



PROJETS  
**R&D**

FINAXYS  
Big Data stack  
for Finance

spécialisation  
banque/  
assurance

expertise  
big data



conseil, projet,  
architecture

2013

**ÉTUDE BIG DATA**

Étude des  
technologies  
Big Data comme  
alternatives crédibles  
dans les SI bancaires

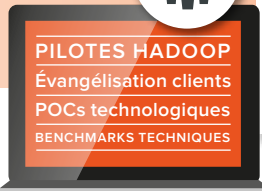


2014



PARTENARIAT  
**ÉDITEUR**

Création de notre offre  
Big Data basée sur la  
solution **SCALED RISK**



2015

CRÉATION  
**DE L'OFFRE**  
conseil Big Data  
FINAXYS

**PRISE DE  
PARTICIPATION**  
AU CAPITAL DE  
**SCALED RISK**

Éditeur Big Data  
pour la Finance



2016

**TALKS  
ET CONFÉRENCES**

Hadoop User Group  
Paris Datageeks  
Distributed Matters  
Elastic Paris



EN 2015

**20  
JOURS**

DE FORMATION  
**BIG DATA**

**7  
WORKSHOPS**

HADOOP  
EN 2015



## NOS EXPERTISES

- HADOOP
- SGOOP
- KAFKA

- HBASE
- ELASTICSEARCH
- HIVE

- STORM
- SPARK
- FLINK

**Alvin RAMGOBEEN**

Group Practice Manager BI & Big Data

+33 (0)1 44 04 50 08

alvin.ramgoheen@gfi.fr

La Porte du Parc - 145 boulevard Victor Hugo 93400  
Saint-Ouen France



## UN MARCHÉ EN MANQUE DE REPÈRES, UNE OFFRE PLÉTHORIQUE, UNE RÉELLE DIFFICULTÉ À DISCERNER LE VRAI

En France, nous observons une lente maturation du marché Big Data avec la difficulté de discerner l'émergence d'un standard ou bien encore d'avoir confiance dans le réel niveau de compétences d'équipes de développement. Au fil des années, nous avons recueilli les marqueurs indéniables favorisant la réussite de projets Big Data. Ceux-ci sont devenus des « invariants » qui représentent le fondement de notre culture et sur lequel nous nous sommes appuyés pour réussir notre transformation interne. Fort de ce renouveau, nous proposons un éventail d'offres dont le scope s'étend de la définition à l'activation de services sur des plateformes industrielles.

**Des bénéfices oui, mais souvent différents.** A l'heure de la bascule des POCs vers les projets en production, beaucoup de projets métiers disposent de ROI. Pour certains, ce ROI se chiffre à plusieurs millions d'euros, mais ces bénéfices sont cependant différents. Par exemple, les gains de parts de marché se rattachent volontiers à la Net Economie, tandis que la réduction des coûts fait partie intégrante du paysage des entreprises plus traditionnelles.

**Les clés de la réussite au démarrage.** Poser un cadre est indispensable. Selon des critères objectifs, il faut définir précisément quels sont les cas d'usages métiers, quel est le niveau d'affinité des ressources internes face à ces nouvelles technologies. Mais il ne faut pas négliger non plus certains critères plus subjectifs comme l'ambition stratégique, la valorisation de l'image ou encore la culture d'entreprise. C'est avec ces derniers qu'il faudra composer, car ces critères représentent les leviers les plus efficaces, une fois bien alignés.

**Lever les freins liés à la culture française.** Contrairement aux projets français, la plupart des projets américains sont « drivés » par le business qui souhaite lancer une offre centrée sur la data. Les développements se basent sur le design de la meilleure architecture technologique à mettre en œuvre par rapport à leur besoin. S'ensuit la réalisation d'un vertical pour un besoin donné qui a pour vocation d'inonder un marché. La réussite est liée au dialogue rapproché entre la direction générale, les métiers, le maître d'œuvre et l'exploitation.

**Plateformes Cloud ou physiques, architecture hybride, distributeur.** Puisque les solutions technologiques évoluent, les questions d'architecture se posent souvent en référence aux pratiques et habitudes des sociétés en termes de choix hardware, de solutions logicielles, de connaissances en langages de programmation ou bien même en méthodologies projets ou en gestion d'environnement. Sur ce volet, nous avons développé une expertise liée à notre expérience sur les impacts en performance. **Bien cadrés, 60 % des projets se font en moins de 6 mois.** Au commen-

cement est le SENS de la donnée et non son caractère « Big ». Une fois ce concept premier bien intégré, il est alors plus facile de se concentrer sur la recherche de la valeur cachée. Il faut s'attacher à redécouvrir son activité et à enrichir progressivement sa vision bâtie sur des a priori. En ce sens, l'architecture hybride conjugue le meilleur de plusieurs univers : celui des bases de données et du Big Data en tirant parti du patrimoine existant, tout en limitant son effort d'intégration. Le respect des méthodes itératives et DevOps restant essentiel.

**Intrication du Big Data, du Digital, de la mobilité, de l'IoT, de la sécurité : une autre nécessité.** Phénomène complexe, s'il en est, le « Data Driven » bouleverse nombre de paradigmes. Les services analytiques se sourcent de masses d'informations de l'IoT ou encore des interactions digitales pour donner naissance à d'autres services. La superposition des calques de données entre eux font émerger des observables qui pour certains seront des leviers d'actions pour les métiers ou des messages directement adressés au client final sur son smartphone...si la CNIL l'autorise. Il est même possible d'envisager l'extension d'un service à l'international via la création d'un tiers de confiance.

**Des services déployés 24/7 sur plusieurs plateformes depuis plus de 5 ans.** La vision holistique apportée par la lecture de l'information à l'échelle d'une activité permet d'en retirer des bénéfices indéniables sur la chaîne logistique, le ressenti client ou l'efficacité des services support, avec pour chacun, le bon niveau d'interface de visualisation. Les ROI dégagés sont atteints d'autant plus facilement que le système apporte des leviers d'actions sur une thématique précise au lieu de supporter un portefeuille de besoins éparpillés et/ou non clairement définis.

**Parier sur le capital humain avant tout :** Les technologies évoluant, il est paradoxalement plus prudent de miser sur les facultés d'adaptation de l'humain. Nous avons mis au point une approche permettant de s'assurer à 80% que nos effectifs retenus sur ces sujets disposent bien du bon jeu de « mindset ». Ceci favorisant vertueusement la réussite des projets, l'évolution des juniors et la progression des seniors. Nous disposons également de moyens conséquents pour nos besoins de R&D, de formations, et pour nos démonstrateurs via des infrastructures hautes performances grâce à notre partenaire HPE.

**Des offres selon les besoins de maturité ou de choix stratégiques :** Nous offrons en plus des prestations de conseil, d'intégration ou de maintenance et d'exploitation un accompagnement basé sur l'écoute et l'appropriation des concepts clés sur ces sujets. Nous proposons également des accélérateurs de détection d'Insights métier via la solution MyDataBall, d'expertise pointue sur notamment le déploiement automatisé de clusters via la société Adaltas. De plus, nous avons élaboré des verticaux sur la Cybersécurité (label France) avec Keenai (VigieSI) ou bien encore sur les Smart Transport ou le Retail.

**Téléchargez notre livre blanc : [www.gfi.world](http://www.gfi.world)**

**3 bénéfices**  
Gain de parts de marché  
Réduction des coûts internes  
Amélioration de l'image de marque

**4 freins majeurs en 2016**  
• Culture d'entreprise  
• Résistance de l'IT  
• Volonté d'autonomie par les métiers  
• Manque de compétences internes

**+ 950**  
Consultants Digital & Data

**+ 1700**  
Panels de projets de l'étude dont 66 % dépendent de la Data !

**+ 300 millions**  
d'objets connectés sur les systèmes en production

**1 solution & un réseau de startups**  
UN VERTICAL MÉTIER EN CYBERSÉCURITÉ Solution de SIEM Keenai (VigieSI)  
• Performance et Sécurité Hadoop  
• Moteur de Dataviz Analytique



# Gfi architecte ensemblier de vos projets Big Data

[gfi.world](http://gfi.world)



**Parce que chaque projet Big Data est l'expression de votre essence et de votre culture Digitale, nous sommes à vos côtés pour relever ces défis avec une passion éclairée et pragmatique.**

Gfi Informatique, dispose de retours d'expériences industriels dans le secteur du Big Data en totalisant plus de 10 plateformes en production pour un total de 2,5 péta-octets de données et 300 millions d'objets connectés. Notre vision du marché se concrétise par des investissements et des prises de positions affirmés qui nous amènent à co-construire les architectures de services orientés data, avec vous et nos partenaires startups. Seul acteur à proposer une offre de services à la pointe du savoir-faire français, allié à l'approche industrielle d'un groupe polyvalent, en gestion des infrastructures, intégration de systèmes et en édition de logiciels, Gfi Informatique opère la réalisation de vos programmes Big Data en tant qu'architecte ensemblier.

Florence LAGET

Directrice Big Data France

+33(0) 820 211 211

bigdata.france@hpe.com

Immeuble Arcs de Seine - 20 quai du Point du Jour 9210  
Boulogne Billancourt France

www.hpe.com




# Hewlett Packard Enterprise

## LE BIG DATA, PILIER DE L'INNOVATION POUR LES ENTREPRISES?

Michael Porter qualifie l'innovation comme étant la clef de la prospérité économique. Dans un contexte économique complexe, exigeant et d'une vitesse jamais vue auparavant, innover semble être la solution pour aider les entreprises à tirer leur épingle du jeu et ainsi prospérer. Parmi les facilitateurs d'innovation, on retrouve le Big Data, souvent qualifié de panacée dans le domaine de la transformation digitale des entreprises. A défaut d'être un remède miracle, le Big Data est surtout un outil puissant pour créer de nouveaux usages et par conséquent, construire le business du futur. Si les mastodontes Google et Apple apparaissent comme les champions de l'innovation dans les classements (BCG 2015), les entreprises plus traditionnelles savent elles aussi que le Big Data représente un véritable levier de croissance. En effet, selon une étude PwC, 76% des entreprises industrielles françaises estiment atteindre un niveau de digitalisation avancé d'ici 2020, contre 38% aujourd'hui où le data analytics jouera un rôle prépondérant (1ère édition de l'étude mondiale de PwC sur l'industrie 4.0).

Si le Big Data est incontestablement un pilier de l'innovation, le véritable enjeu se porte sur son utilisation pour servir au mieux la stratégie de l'entreprise.

Les start-up ont un avantage sur les grandes entreprises car elles sont digitales dès leur création et plus agiles pour intégrer le Big Data dans leur business model. Deux licornes françaises, BlaBacar et Criteo se sont développées avec cette approche dans le but de mieux prédire et répondre aux besoins de leurs clients et prospects. Ainsi, BlaBlaCar cherche à comprendre le comportement des utilisateurs pour pouvoir développer ses services. Illustration parmi d'autres, l'analyse des données a permis d'identifier un besoin spécifique à partir de données non structurées : certaines passagères préfèrent que la conductrice soit également une femme. Dans un autre secteur d'activité, Criteo utilise les outils HPE pour répondre en temps réel à des requêtes portant sur des milliards de lignes afin de générer des publicités personnalisées sur internet. Chez Uber, la combinaison des outils HPE Big Data et machine learning permet à la société de services de transport d'analyser et comprendre le comportement, les lieux et les préférences de ses clients afin de gérer plus efficacement la disponibilité et le positionnement des chauffeurs. Mieux encore, certains facteurs comme le jour du versement des allocations chômage et des aides sociales expliquent certains pics de commandes. Les entreprises plus traditionnelles ne sont pas en reste : grâce aux solutions HPE Big Data, la Société Générale peut désormais répondre aux besoins de ses directions marketing et commerciale grâce à l'analyse du parcours de ses clients (et par conséquent de leurs centres d'intérêt) sur les canaux web et mobiles. Par exemple, connaître les segments de clients qui s'intéressent à l'assurance ou aux crédits, les modes préférés par certains clients entre le web, la mobilité ou l'agence, ou encore vérifier l'efficacité des campagnes en ligne pour mesurer l'impact sur les pages vues associées à ces contenus.

Dans le domaine sportif, l'innovation est, elle aussi, essentielle pour se démarquer, aller plus loin et gagner les précieuses dixièmes de secondes qui feront la différence. Dans la version moderne de la course automobile, la Formule E fait courir des véhicules propulsés par un moteur électrique. Là encore le Big Data est à l'honneur, au cœur même de cette avancée technologique : grâce aux capteurs audio et vidéo intégrés, les données sont analysées en temps réel et les équipes peuvent s'ajuster en permanence pour gagner en compétitivité.

Mais l'innovation n'est pas uniquement synonyme de croissance économique. Elle est aussi porteuse d'une dimension civique lorsqu'elle est au service de la sécurité. En Espagne, le Ministère de l'intérieur a pour mission de lutter contre le terrorisme et le crime organisé. Grâce à la stratégie Big Data, les forces de sécurité intérieures peuvent traiter très rapidement ce volume considérable de données pour analyser les comportements et détecter à temps les menaces potentielles qui pourraient mettre en péril les concitoyens. Et ainsi, sauver des vies.

L'innovation est partout, les domaines d'applications sont encore vastes, notamment dans le secteur de la santé, de l'énergie, de l'agriculture. La quatrième révolution industrielle est en marche et le Big Data sera un pilier fondamental pour accompagner les entreprises, privées ou publiques, dans leur transformation digitale.

# Accélérer l'intelligence des objets

Hewlett Packard Enterprise et Intel valorisent  
l'intelligence des objets à travers une nouvelle  
solution analysant la donnée à tout moment  
et en tout lieu.

WEATHER  
ALERT



Pour en savoir plus, visitez [hpe.to/loT](http://hpe.to/loT)

Accélérer l'avenir



**Hewlett Packard  
Enterprise**



## L'ANALYSE DE LOGS COMME CAS D'USAGE IDEAL

Hurence accompagne ses clients dans la réalisation d'infrastructures Big Data et de leurs cas d'usages, du projet pilote jusqu'à sa mise en production.

Avec des années d'expérience, nous savons que le premier challenge d'une initiative Big Data est de trouver le bon angle de travail. Celui qui permettra à notre client de se former sur ces technologies nouvelles, aux équipes métiers d'adhérer à l'initiative parce qu'il représente à la fois une vision et une innovation, et enfin celui qui va convaincre le management du nécessaire retour sur investissement. Nous partons alors à la recherche du "cas d'usage idéal", autant dire un véritable « Graal ».

Nous avons accompagné nombre de grandes sociétés à la recherche de ces "cas d'usage" et avons développé avec l'expérience une méthodologie simple et adaptative qui va à l'essentiel. Bien que le Big Data aille de paire avec l'initiative de transformation digitale "stratégique" de l'entreprise, le prendre sous cet angle c'est se paralyser dans une approche focalisée sur la réflexion et à plus long terme. C'est se priver du "mouvement", de l'"agilité" que doit représenter une initiative Big Data. C'est ignorer que les idées fusent en pratiquant... et que les idées les plus simples, bien traitées, bien creusées font les grandes réussites, les « success stories ». Fort de ces constats, Hurence prône donc de rompre avec les modèles classiques et opte pour toutes les dimensions d'un projet Big Data, sur une méthodologie agile qui permet d'avancer en faisant.

Un projet Big Data est un projet de « data engineering » à savoir un projet qui allie développement et science de la donnée. Ce sont des projets dont la complexité peut effrayer mais Hurence prend en charge les zones complexes de ces projets et fait monter en compétences l'équipe client grâce à ses formations (à ce jour une vingtaine de modules) pour qu'elle en maîtrise tous les autres aspects.

Notre accompagnement refuse les initiatives "lourdes" dans lesquelles on peut être amené à convoquer des ateliers "métiers" sur des mois pour souvent accoucher d'un « cas d'usage consensus » en général décevant. Nous encourageons nos clients à revenir sur des idées simples et rapides à mettre en œuvre. Notre méthodologie permet de dégager rapidement quelques unes de ces idées qui semblent a priori peu révolutionnaires, mais si traitées de manière innovante ouvriront de nouvelles et belles perspectives.

Dans cet esprit, l'un des plus beaux accompagnements que Hurence ait réalisés, sachant que des évolutions sont prévues, est celui de Carrefour dans la conception et la mise en place du traitement des tickets de caisse et la gestion des stocks en temps réel (<http://www.lemagit.fr/etude/Carrefour-dote-son-SI-dune-architecture-data-centric>). Ces travaux ont été réalisés avec l'équipe de la DSI BI/Big Data, une équipe incroyablement ouverte au changement et à l'innovation, réceptive à de nouvelles méthodes d'organisation et de développement.

En fait si le traitement de tickets de caisse ou plus généralement le traitement de logs (des tickets de caisse ne sont rien d'autre que des mini logs d'événements élémentaires) vous semble une idée bateau, en fait elle ne l'est pas...

Lorsque l'on met en place une analyse de logs temps réel dans son entreprise, on pose, en réalité les bases de ce que sera l'entreprise de demain. Une entreprise connectée au monde, ou tout est événement, un événement que l'on sait capter, analyser et sur lequel on saura réagir immédiatement.

On transforme ainsi son « Data Lake », un artefact statique par nature, en « Data River » une structure dynamique, réactive où les données peuvent aussi être en mouvement. De facto, on se prépare à l'Internet of Things (IoT) et à sa révolution : des milliards d'événements seconde pouvant être déversés dans la rivière. Dans tous les cas, on pose les bases d'un socle qui de toute façon sera celui du futur.

Factorisant l'expérience de nombreux accompagnements sur l'analyse de logs ou de données apparentées (tickets de caisse, données de capteurs industriels), Hurence a publié en logiciel libre log-island (<https://github.com/Hurence/logisland>), démontrant ainsi son implication et sa présence dans la communauté du monde libre.

Log-island (<https://github.com/Hurence/logisland>) est une solution avancée de traitement et d'analyse de données temps réel basée sur les technologies les plus performantes de l'écosystème Big Data pour l'analyse de volumes de donnée par unité de temps considérables (Nifi pour acheminement des données brutes, Kafka comme bus de messages à grande bande passante, Spark pour l'exécution parallélisée des processeurs d'analyse événementielle). Outre le packaging simplifié de cette infrastructure d'« Event Mining » à grande échelle, Hurence a également offre un certain nombre de processeurs directement utilisables pour traiter des logs classiques (syslogs, access logs, etc.), détecter des "outliers" (phénomènes hors-norme) sur des séries temporelles type IoT, lever des alertes sur des règles métier. Nos clients peuvent alors capitaliser sur des éléments communs et se focaliser sur ce qui leur est spécifique: leurs propres règles métier, en écrivant facilement leurs propres processeurs temps-réel sans se préoccuper des problématiques complexes du traitement parallèle

Mettez vos données en mouvement  
du "Data Lake" vers la "Data River"



**HURENCE**

Vos experts Big Data  
<http://www.hurence.com>  
[contact@hurence.com](mailto:contact@hurence.com)



## IDEATRANS, AGENCE DE TRANSFORMATION DATA DIGITALE & INNOVATION !

### Promesse

Votre performance ! iDeaTrans vous aide à développer et transformer vos métiers et votre business par la Data, le Digital et l'Innovation :

- **Révéler et clarifier votre maturité Data** : place de la Data dans mon organisation ? stratégie et enjeux Data ? ...
- **Sourcer vos besoins Data** : quelles Data utiles ? interne ? externe ? ...
- **Faire émerger vos opportunités Data** : conditions d'émergence ? organisation ? marché ? besoins ↔ Data ? ...
- **Créer de la valeur dans vos métiers et au-delà** : efficacité ? KPIs ? analytics ? nouveaux paradigmes ? ...
- **Valoriser la Data en toute sécurité** : enrichir ? mettre en avant ? protection des Data personnelles ?

### Marchés

iDeaTrans s'engage avec et envers les Organisations qui cherchent à répondre à une volonté et des problématiques de Transformation ou d'Innovation (améliorer ou réinventer), qu'elles soient métiers, organisationnelles ou techniques, en s'appuyant sur toutes les ressources potentielles (360° interne - externe, Data et Digitale), et sur une méthode reconnue (Label d'OR) d'Innovation ouverte.

### Clients

Laboratoire X (direction technique d'un projet de transformation Digitale Monde), Monoprix (créer le schéma directeur Data, Référentiels User Centric, Omni Canal), QLIK (accompagner le challenge Open Data France), BEL (analyser le marché de la Data Marketing Fromage en Algérie), IPSEN (révéler la Data Digital dans la e-Santé, Start-Ups), AGORA (créer de la Valeur Data au sein de communautés), C3 COP21 (organiser et utiliser l'Open Data EDF-ERDF-Météo France-IGN-CNES-IRSTEA-ETALAB-MEDDE pour instruire la problématique Climat et faire émerger 29 projets dans toutes la France), GfK IFR (créer et gouverner la Data & Digital SI pour développer le Business Clients et la Transformation du Groupe, diriger les Opérations monde IFR), UBISOFT - LECLERC (Conseil autour des référentiels MDM), POLE EMPLOI (révéler et clarifier la maturité Data afin d'initier un projet Big Data), MNMS (contribuer à faire émerger des solutions innovantes dans les organisations (EDF, AG2R La Mondiale, SYNERTHON ATM Toulouse,)), de l'idée au marché, etc.

### Savoir-faire

+25 années d'expertise dans la valorisation de la Data (Marketing, Retail, Open Big Smart Data), identification, transformation et création, de la source au client, des contenus aux usages métiers. +20 ans de transformation des organisations en s'appuyant sur le Digital-Data, l'Innovation Métiers (Plateformes ETL, MDM, CRM, DWH, BI BigData, LMS e-Learning, etc). Accompagnement de Start-Ups depuis 1997 (conseil & investissement, de l'idée au marché).

### Contributions

AFNOR, AFDEL TECHIN, IDC, Gartner, CIO, CIOnet, IT for Business, 01, IT Expert, Mag IT, CIGREF, IBM, Oracle, Informatica, QLIK, Business Objects, AFAl, AFCDP, Finance Innovation, EBG, Alliance Big Data, Big Data Paris, Cap Digital, ESSEC, Novancia, AFM, CMiT, Distributique, Printemps des Etudes, G9+, Total, Micropole, Neurones, B&D, Linkbynet, etc...

### Organisation innovante

iDeaTrans réunit les acteurs nécessaires et des experts indépendants depuis +15 ans, organisation originale ouverte, et collaborative. iDeaTrans a choisi la méthode et démarche d'innovation MyNeedMySolution (MNMS s'appuyant sur l'intelligence collective depuis + 15 ans) pour révéler, faire émerger les idées et besoins, et les transforme en solutions concrètes et pérennes.

### Fondateur CEO

Fabrice Benaut (Paris 2), un activiste du numérique, révélateur et accélérateur de la transformation et de l'innovation, fondateur de la société iDeaTrans, Fabrice est aussi associé dans MNMS et dans plusieurs start-ups. Intra-entrepreneur innovant, Fabrice a été Directeur des Systèmes d'Information et Opérations, en charge des Projets Métiers, membre du COMEX de GfK IFR Global (60 pays), et CIL. Ses 25 années d'expérience internationale dans le Business de la Data (360° métiers, Marketing, Marché, Distribution, Usages, Comportements, Traces, etc) font de lui un expert en matière de Big Data et de Transformation (numérique) de l'entreprise (startup et grands groupes). Fabrice a été également Commissaire aux Données pour le [www.c3challenge.com](http://www.c3challenge.com) COP21.

Proche du monde des start-ups depuis 1997 (co-working) où il intervient en qualité de conseil et d'investisseur (Privowny, Stample, PCB xField Paintball, Citymeo, Integra, Keakr, etc...), Fabrice anime des ateliers « de l'idée au marché » et de co-design (Focus Group) avec MNMS.

Conférencier reconnu, Fabrice Benaut est président de l'Alliance Active Data, et membre -très - actif dans plusieurs associations regroupant le monde du Digital Data Marketing, DSI-CIO et Business, professionnels du monde numérique. Il aime partager et enrichir son savoir Data Digital en intervenant aussi régulièrement dans les écoles, les universités et les groupements professionnels. Fabrice est aussi Expert Judiciaire Eco-Fi Etudes de Marché Data (Versailles).





OÙ SONT LES DONNÉES ?

POUR QUELLE  
PERFORMANCE METIER ?



Valoriser  
la donnée en  
toute sécurité

iDeaTrans VOUS AIDE A DEVELOPPER ET TRANSFORMER  
VOS METIERS ET VOTRE BUSINESS PAR LA DONNEE

Révéler,  
clarifier votre  
"maturité données"

Sourcer  
vos besoins  
données  
(open, big data,  
interne, externe)

Créer  
de la valeur  
dans vos métiers  
et au-delà

Faire émerger  
vos opportunités  
données



iDeaTrans • agence de transformation data & digitale • Paris  
fabrice.benaut@ideatrans.net • www.ideatrans.net

Partenaires



MyNeedMySolution  
Agence d'innovation



Alliance Active Data  
Première communauté  
du Big Data



# KALANE

**vincent GRISON**  
Associé Fondateur

+33(0)1 71 18 39 01  
contact@kalane.com  
9 rue Jean Mermoz 75008 Paris France

CONTACT



www.kalane.com



## STRATEGIE. DIGITAL. DATA. CONSULTING

Depuis plus de 10 ans, kalane consulting, cabinet de conseil en stratégie, accompagne ses clients en France et à l'international sur leurs projets stratégiques : développement & diversification, transformation & performance digitale.

Les expertises digitales et data du cabinet, alliées à une approche agile et pragmatique de la stratégie, ont permis à kalane consulting d'accompagner plus de 50 clients sur une dizaine de secteurs d'activité : médias, communication, distribution, assurance, automobile, produits manufacturés, nouvelles technologies, e-commerce, grande consommation.

## UNE APPROCHE MIXANT CONSEIL, DATA ET LEAN STARTUP

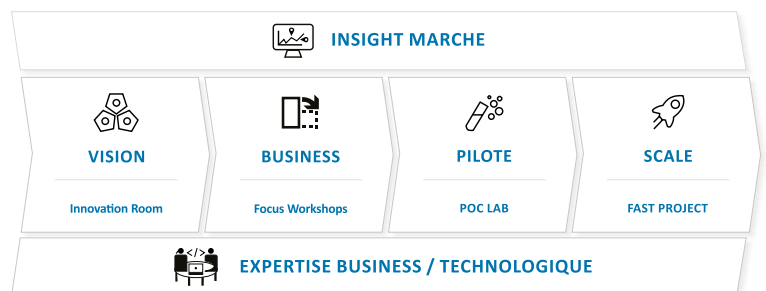
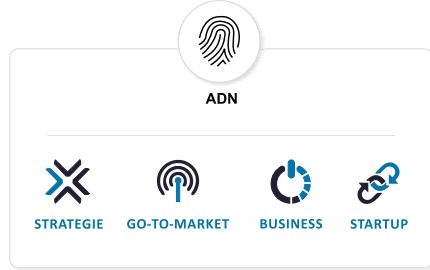
- Le cabinet kalane consulting se distingue par une approche associant :
- La compréhension des enjeux stratégiques et des modèles économiques des clients
  - La data analyse, data exploration et data science pour transformer des intuitions en convictions
  - La capacité à accompagner les clients de l'innovation jusqu'au go-to-market
  - Des méthodologies pragmatiques et agiles inspirées du lean startup : concentration, itérations, pivot et mesure

## UNE OFFRE DE CONSEIL « BUSINESS CENTRIC »

L'offre de conseil de kalane intègre le digital et la data dans une réflexion et une compréhension globales du business de ses clients. Pour conduire avec succès les missions de conseil, les consultants kalane sont seniors, issus de parcours mixte conseil & startup, passionnés par la data, le digital et les technologies, ... et totalement imprégnés des modèles économiques des industries que le cabinet adresse.

## UNE DEMARCHE DATA END-TO-END

La DATA est intégrée au sein du cabinet kalane consulting dans une approche globale permettant de définir une ambition et cerner les enjeux stratégiques tout en exécutant des use cases métiers dans une approche expérimentale. Cette méthodologie permet à nos clients d'investir de façon mesurée dans un premier temps pour conserver toutes leurs capacités d'investissement lorsque les convictions sont établies.





## STRATEGIE DATA & CRM – GROUPE MEDIA LEADER EN EUROPE

Pour un des leaders des médias européens, le cabinet kalane consulting a contribué à la définition d'une stratégie data globale pour l'ensemble des business du groupe. L'ensemble de l'industrie des médias fait face à un contexte de concurrence des plateformes internationales et un mouvement user centric du marketing et de la communication. Kalane consulting a co-construit avec son client une stratégie et une feuille de route opérationnelle pour exécuter les premiers use cases data : publicitaire, éditorial, e-commerce. Les travaux réalisés au cours de cette mission ont été les suivants :

- Analyse des tendances du marché et benchmark des initiatives des groupes médias à l'international
- Analyse de l'ensemble des assets data de l'ensemble des directions et filiales du groupe
- Définition et priorisation des use cases pour les différents business du groupe
- Définition d'une feuille de route sur les technologies et les expertises nécessaires
- Sourcing, sélection d'une data management plateforme
- Setup d'un data lake pour exécuter les premiers use cases

## PERSONALISATION ET NOUVEAUX BUSINESS DATA – PUR PLAYER E-COMMERCE

Pour un acteur majeur de l'e-commerce européen, kalane consulting a conduit une mission ayant pour objet d'utiliser une approche data centric pour personnaliser l'expérience client d'une part et créer une nouvelle ligne de business d'autre part. Une approche agile sur la base d'exécution de case en mode pilote a été déployée :

- Co-construction de use cases sur la personnalisation et le new business data
- Recommandation sur la stratégie technologique sur la data
- Setup d'une architecture de data expérimentation parallèle de l'architecture de production
- Ingestion de l'ensemble des données pertinentes au sein d'un data lake : catalogue, ventes, CRM, usages digitaux
- Calculs de segments pertinents dans un data lake pour la personnalisation de l'expérience et la commercialisation auprès d'annonceurs

## STRATEGIE ET GO-TO-MARKET POUR UNE ALLIANCE DATA

Pour des groupes leaders dans leur secteur, kalane consulting a réalisé une mission visant à étudier les opportunités de synergies autour de la data. La data était identifiée dans ce contexte comme un levier de performance business mais aussi un levier de transformation interne des processus et des métiers. Le cabinet a accompagné les groupes sur un modèle de partenariat avec les phases suivantes :

- Identification des synergies et des risques associés à ce projet
- Analyse des assets data respectifs des groupes
- Identification de use cases permettant de dégager de la valeur additionnelle
- Setup d'un data lake et exécution des premiers use cases
- Construction de cluster data, algorithmes de lookalike et solutions de prédiction
- Définition du modèle de gouvernance entre les groupes
- Définition du modèle de gouvernance sur les données

## STRATEGIE DATA POUR L'ACHAT MEDIA - INDUSTRIE AUTOMOBILE

Pour un constructeur automobile mondial, l'activation de sa data first party dans les transactions publicitaires sur le digital constitue un enjeu majeur pour l'efficacité de la conquête de nouveaux clients. Fort de son expertise de la data, de la publicité et de l'industrie automobile, kalane consulting a étudié l'opportunité de mettre en oeuvre une data management plateforme permettant d'améliorer le ROI publicitaire des campagnes digitales – ciblage, personnalisation du message publicitaire,... - sur les principaux leviers de communication du marché chinois.

- Analyse de l'écosystème publicitaire et des leviers de communication pertinents à activer avec de la data
- Analyse des sources de données disponibles au sein des plateformes chinoises
- Identification des solutions technologiques et des partenariats data pertinents en Chine
- Construction du business plan : coûts des solutions, budgets médias impactés, bénéfices attendues sur les ventes
- Sélection des solutions et partenaires technologiques locaux
- Elaboration de l'organisation et de la gouvernance data entre les différentes entités du constructeur et ses agences



# KEYRUS

CONTACT

## KEYRUS, CRÉATEUR DE VALEUR À L'ÈRE DE LA DATA ET DU DIGITAL

Acteur international du conseil et des technologies, spécialiste de la Data et du Digital, **Keyrus** a pour mission d'aider les entreprises à tirer profit du paradigme de la Donnée et du Numérique pour accroître leur performance, faciliter et accélérer leur transformation et générer de nouveaux leviers de croissance, de compétitivité et de pérennité. Plaçant l'innovation au cœur de sa stratégie depuis 20 ans, **Keyrus** développe une proposition de valeur unique sur le marché autour d'une offre novatrice qui s'appuie sur la combinaison de trois expertises majeures et convergentes :



### MANAGEMENT & TRANSFORMATION

Stratégie & Innovation | Pilotage de la Performance  
Transformation Digitale | Accompagnement des Projets



### DATA INTELLIGENCE

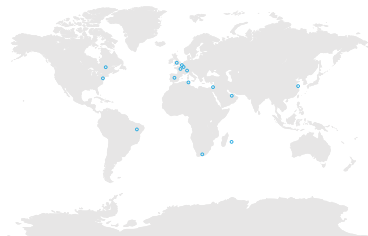
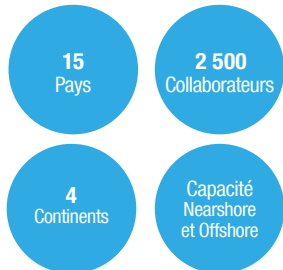
Data Science | Big Data Analytics | EIM  
Business Intelligence | CPM/EPM



### DIGITAL EXPERIENCE

Innovation & Stratégie Digitale | Marketing Digital & CRM  
Commerce Digital | Performance Digitale | User Experience

### EFFECTIF ET IMPLÉMENTATION :



FRANCE - AFRIQUE DU SUD - BELGIQUE - BRÉSIL - CANADA - CHINE - ÉMIRATS ARABES UNIS  
ESPAGNE - ILE MAURICE - ISRAËL - LUXEMBOURG - ROYAUME UNI - SUISSE - TUNISIE - USA

## ACTEUR HISTORIQUE DE LA BI, SPÉCIALISÉ DANS LE BIG DATA

**Keyrus** fait partie des acteurs historiques du marché de la Business Intelligence sur lequel elle a su développer des expertises multiples liées à la gestion et l'exploitation de la donnée. L'ESN a été l'une des pionnières du secteur à se développer sur les activités liées au « Big Data ». Acteur majeur et moteur de cet univers, **Keyrus** a mis en place des actions et des partenariats avec de prestigieuses écoles afin de former et recruter les ingénieurs Big Data de demain, notamment à travers l'organisation plusieurs fois par an de Boot Camps. **Keyrus** accompagne aujourd'hui en France et à l'international les entreprises pour les aider à tirer profit des enjeux liés au Big Data à travers une offre d'accompagnement personnalisée :

- > CONDUITE DE POC
- > INDUSTRIALISATION BIG DATA
- > FORMATIONS ET SÉMINAIRES
- > STRATÉGIE ET FEUILLE DE ROUTE
- > ACCOMPAGNEMENT MÉTIERS & DATA

## DÉCOUVREZ RAYS, LA NOUVELLE PLATE-FORME CLOUD QUI ACCÉLÈRE VOS PROJETS BIG DATA ET ANALYTICS

RAYS est une plate-forme flexible qui valorise des données massives et multistructurées à travers l'utilisation d'une bibliothèque de modules de Data Intelligence. Développée par **Keyrus**, RAYS vous permet de gérer vos projets Big Data, depuis la récolte des données jusqu'à la visualisation en passant par une modélisation algorithmique. Basée sur l'expérience et le savoir-faire de **Keyrus** dans la Data, RAYS offre une large bibliothèque d'Apps implémentées, mises à votre disposition et vous permettant de vous concentrer sur le cœur de vos projets Big Data. Véritable accélérateur technologique, profitez du Cloud et du déploiement automatique des outils Big Data qui vous simplifient l'accès à l'analyse de vos données.

## NOS RÉFÉRENCES DANS LE BIG DATA

### Création d'un modèle prédictif de contenu

- En seulement trois mois, **Keyrus** a répondu aux besoins d'un quotidien gratuit destiné aux usagers de transports en commun en France : développer la consultation en ligne des articles et accroître le trafic sur son site internet. L'objectif était de construire un modèle prédictif dans un contexte évoluant rapidement et suivant des effets de masse difficilement identifiables. Pour cela, **Keyrus** a travaillé avec des Reader logs pour livrer une analyse complète du trafic par canal et du comportement de l'utilisateur. Ces analyses ont mené à une segmentation des comportements pour définir la meilleure approche et à l'installation de l'outil «Recommandation de nouvelles» qui a permis d'augmenter le trafic et le temps passé sur les articles du site mais également d'optimiser les contenus en fonction du profil utilisateur et d'améliorer la stratégie publicitaire.

### Construction de tableaux de bord aux sources multiples

- Les experts Big Data de **Keyrus** ont répondu aux attentes de l'un des leaders mondiaux des brasseurs qui désirait se rapprocher de ses consommateurs afin de mieux comprendre le parcours client, leurs opinions et leurs préférences. L'objectif était de conduire l'innovation produit en s'appuyant sur la connaissance du client et des analyses Big Data. Pour cela, **Keyrus** a mis en place des tableaux de bord et des rapports d'analyses pour aider aux opérations journalières. Collecter, structurer et analyser les enquêtes externes a permis de mieux définir les habitudes de consommation vis-à-vis des boissons alcoolisées et non alcoolisées. De plus, l'analyse des opinions sur les réseaux sociaux, blogs et autres médias en ligne a aidé à obtenir la vision globale des attentes du consommateur en associant différentes sources de données. **Keyrus** a déployé chez ce client un tableau de bord intuitif facilitant l'appropriation des analyses et a permis de conduire l'amélioration et l'innovation produit tout en assurant une meilleure réponse aux attentes des consommateurs.

### Exploitation de flux de données générées par des compteurs intelligents

- L'un des leaders français de la gestion de l'eau, des déchets et de l'énergie a fait appel à l'expertise de **Keyrus**. Afin d'augmenter la fidélité de ses clients en développant des compteurs intelligents. **Keyrus** a développé un moyen de gestion des flux de données générés par les compteurs intelligents grâce à l'association de son système PaaS et d'un modèle performant utilisant des algorithmes d'apprentissage automatique. L'objectif était la création d'une application sur Smartphone permettant au consommateur de gérer sa consommation d'énergie et d'eau en temps réel. Le modèle combine les données utilisateurs, sa consommation passée, celle du voisinage ainsi que la consommation de sa catégorie. Ces actions combinées ont abouti à une augmentation significative de la fidélité client.

## LE BIG DATA

### UNE RÉALITÉ À VOTRE PORTÉE !

LEADER DU BIG DATA, KEYRUS POSSÈDE À LA FOIS LA CONNAISSANCE MÉTIER, L'EXPERTISE ANALYTIQUE ET LA MAÎTRISE TECHNOLOGIQUE INDISPENSABLES AU SUCCÈS DE VOS PROJETS BIG DATA.

- > Valoriser les données massives en les transformant en connaissances exploitables
- > Augmenter les capacités d'analyse pour une meilleure anticipation et prédiction
- > Améliorer la performance des dispositifs digitaux de nouvelle génération
- > Optimiser et faire évoluer les architectures décisionnelles

## KEYRUS VOUS ACCOMPAGNE SUR L'ENSEMBLE DE VOS PROJETS BIG DATA ET ANALYTICS

· MARKETING PERSONNALISÉ

· MARKETING TEMPS RÉEL ET UBIQUITAIRE

· CONNAISSANCE CLIENT 360°

· PRICING DYNAMIQUE

· LUTTE CONTRE LA FRAUDE

· GEOFENCING

· MAINTENANCE PRÉDICTIVE

· USINE DU FUTUR

· SMART CITIES  
SMART GRID

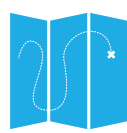
· IOT / IDO

· MTOM



### FORMATIONS ET SÉMINAIRES BIG DATA

- > Préparation et organisation de séminaires d'initiation :
  - métiers
  - technologique
  - analytique
  - juridique
- > Formations d'approfondissement ciblées sur ces thématiques



### STRATÉGIE ET FEUILLE DE ROUTE BIG DATA

- > Analyse et priorisation des cas d'usages
- > Feuille de route sur plusieurs dimensions :
  - cas d'usages
  - données
  - technologies
  - compétences
  - organisation...



### CONDUITE DE POC BIG DATA

- > Analyse du potentiel de nouvelles données
- > Mise en œuvre de modèles analytiques (Machine Learning...)
- > Démonstrateur technologique...



### ACCOMPAGNEMENT MÉTIERS & DATA

- > Structuration et montée en compétence des équipes
- > Accompagnement de projets par des compétences analytiques spécialisées
- > Formations analytiques :
  - R
  - Python
  - Spark...



### INDUSTRIALISATION BIG DATA

- > Définition et mise en œuvre d'architectures, de plate-formes et d'organisations Big Data
- > Mise en œuvre de projets Big Data :
  - Portage technologique de POC
  - Durcissement d'un prototype
  - Évolution d'une solution existante

# MAPR TECHNOLOGIES SAS

**Nicolas POULAKAKIS**

Regional Sales Director

+33(0)6 87 98 11 90

npoulakakis@mapr.com

47 Boulevard Georges Clemenceau 92400 Courbevoie  
France



www.mapr.com



## HISTORIQUE

Société créée en 2009, à San José – Californie. La société compte, début 2016, près de 450 collaborateurs répartis dans plus de 20 pays. MapR compte plus de 700 clients à travers le monde utilisant la MapR Converged data platform pour répondre à leurs problématiques Big Data.

## CŒUR D'ACTIVITE

MapR propose le meilleur des technologies Big Data (Hadoop, Spark, NoSql, Streams, SQL Ansi) dans une seule et unique plateforme : MapR Converged Data Platform.

MapR aide ses clients à répondre à des problématiques business telles que la vision 360° client, la détection de fraude, la gestion de risque, besoins de personnalisation et recommandation, Analyse prédictive, ou les Objets connectés (IoT) pour ne citer que les principaux cas d'usage.

## PRESTATIONS/PRODUITS/SERVICES PROPOSES

MapR Converged Data Platform est disponible dans sa version communautaire gratuite, ainsi que la version Enterprise. Cette dernière offre, en plus d'un support professionnel 24x7, toutes les garanties de performance, de disponibilité, de sécurité et redondance nécessaire à une mise en production d'applications Big Data. En outre, MapR propose des services de consulting ainsi que des formations en ligne gratuites ou en sessions de groupe intra entreprises.

## SECTEURS CIBLES

Les marchés cibles sont principalement les Services Financiers (Banques de détail, Banques d'investissement, Assurances), le Retail, l'Industrie, les Médias, les Télécom, le service public et bien évidemment, le secteur des nouvelles technologies.

## COUVERTURE GEOGRAPHIQUE

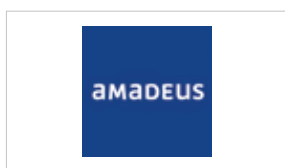
MapR est présent dans plus de 20 pays, répartis sur 3 continents majeurs, les Etats-Unis, l'Europe et l'Asie-Pacifique. Cette répartition permet à MapR d'être au plus près de ses clients et leur fournir ainsi le meilleur service en terme de conseil, assistance consulting et bien évidemment, un support 24x7x365 « follow the sun ».

## AVANTAGES CONCURRENTIELS

Historiquement, MapR est reconnu pour fournir la seule distribution Hadoop hautement disponible, dédiée aux application critiques d'entreprise. Outre cette proposition de valeur fondamentale pour les clients, MapR a accentué ses efforts d'ingénierie pour réduire la complexité d'intégration des nombreux composants d'une plateforme Big Data, à savoir Hadoop, Spark, le NoSQL, le Streaming et les moteurs de requête SQL. En faisant converger toutes ces technologies dans une seule et unique plateforme, MapR réalise ainsi un tour de force en simplifiant l'intégration de ces composants essentiels tout en garantissant la haute disponibilité et la sécurité de l'ensemble. Se positionnant bien au delà d'Hadoop, MapR offre ainsi à ses clients la seule plateforme Big Data d'entreprise taillée pour porter les projets critiques et sensibles de ses clients.

Plus de 700 clients prestigieux, tous secteurs d'activités confondus, font confiance à MapR et bâtissent, ensemble, leur Data Platform de demain.

## PARMIS NOS CLIENTS



# The MapR Converged Data Platform Drives Innovation

MAPR<sup>®</sup>



# WHAT ARE YOU DRIVING?

**MapR Technologies SAS**

47 Boulevard Georges Clemenceau | 92400 Courbevoie | France

[sales-southeu@mapr.com](mailto:sales-southeu@mapr.com)

[www.mapr.com](http://www.mapr.com)

**Laurent VIDAL**  
Country Director

+33 (0)1 53 53 67 84  
Laurent.Vidal@marklogic.com  
23 - rue Balzac 75008 Paris France



## HANNOVER RE, AVEC UN VOLUME D'ENVIRON 17 MILLIARDS D'EUROS, EST LE TROISIÈME RÉASSUREUR DU MONDE. HANNOVER RE GÈRE LES TRANSACTIONS DE TOUTES SORTES DE RÉASSURANCES (immobilier, accidents, assurance-vie et santé).

### Défis

Hannover Re propose, intégrée à son portefeuille de services de réassurance, la nouvelle génération de solutions de souscriptions automatisées avec hr | ReFlex, une combinaison innovante de systèmes de points de vente et d'évaluation des risques spécialement conçus pour les assureurs cherchant à déployer des produits d'assurance entièrement numériques ou à ajouter des canaux en ligne et mobiles pour leurs ventes. La société voulait ajouter une dimension « Informations de gestion » pour aider ses clients à accéder à leurs données pertinentes avec une vue à 360 degrés afin d'analyser les risques et prendre les décisions adéquates. Plus spécifiquement, Hannover Re avait besoin d'une infrastructure capable :

### D'intégrer et de gérer des volumes massifs de données hétérogènes

L'équipe a vite compris qu'une base de données relationnelle ne pouvait pas servir de fondation pour un nouveau système de souscriptions automatisé, en raison de la nature des données. Les données provenant de points de vente sont hétérogènes (chaque client utilisant des procédures de vente et des approches de souscription différentes), et le contenu non structuré tel que les rapports de laboratoires et les images scannées doit être mis à la disposition des décisionnaires. En outre, ces enregistrements sont riches et complexes, avec parfois des milliers d'aspects de données importants intégrés.

### De réduire le fardeau administratif à l'aide d'un système entièrement intégré

Hannover Re se devait de choisir une plate-forme pouvant fournir toutes les fonctionnalités nécessaires de manière cohérente sans générer de complexité opérationnelle. L'objectif consistait à tirer profit des requêtes structurelles et de fonctions linguistiques et en texte intégral sans se soucier de la cohérence des index et de l'organisation de tous les services connexes impliqués. Des solutions concurrentes proposaient des combinaisons de composants indépendants (par exemple, une base de données et un moteur de recherche), ce qui se serait traduit par une complexité accrue en termes de développement et d'exploitation.

### De proposer une expérience utilisateur d'exception

Une demande remplie sur un point de vente est relativement vague, avec des milliers de points de données pertinents, contenant des informations telles que l'état de santé et les antécédents. Tous ces points de données doivent être entièrement disponibles via une recherche intégrée (structurelle et en texte intégral) afin d'analyser les profils à risque, de détecter les schémas à risque précédemment masqués et de proposer des options aux clients pour améliorer leur exploitation commerciale.

### De réduire les risques

Les données étant liées à des problèmes de santé sensibles, la sécurité inquiétait énormément Hannover Re. L'équipe a évalué une solution open source, mais a considéré que le niveau de sécurité était insuffisante.

### La solution

La base de données MarkLogic sert de sous-couche de base de données et de stockage pour une application analytique faisant partie de la suite de produits hr | ReFlex product. L'application analytique basée sur le cloud permet d'obtenir une vue à 360 degrés de toutes les informations pertinentes ; elle est fournie dans un SaaS à des clients dans le monde entier.

Le composant MI dispose de fonctions analytiques supérieures pour les données client collectées dans un entrepôt de données centralisé. Pour chaque installation hr | ReFlex déployée chez un client, MI collecte les données provenant d'applications d'assurance lors de leur cheminement par les différentes étapes du processus numérique de vente d'as-

surances. Des informations sont enregistrées à chaque étape (ex. : état de santé d'un demandeur), et peuvent être analysées pour améliorer les profils à risque agrégés enregistrés par l'assureur et le réassureur. La solution permet d'analyser tous les détails soumis pendant le processus de demande en ligne, notamment les détails de santé déclarés par un demandeur d'assurance-vie.

### Expérience client améliorée

Les nouvelles fonctions analytiques de Hannover Re ont été accueillies avec satisfaction par leurs clients, car elles permettent de mieux cerner les demandes et les plans et de prendre ainsi des décisions prenant mieux en compte les risques, ainsi que d'obtenir les commentaires des clients lors de leur cheminement.

### Gestion des risques

Avec une vue à 360 degrés de toutes les données, les clients peuvent analyser les données pour améliorer leur compréhension de la nature des contrats et des risques afférents. En outre, les fonctions de sécurité de la base de données MarkLogic (anonymat des fragments de données, sécurité élevée de la couche de transport et mécanismes robustes d'authentification et d'autorisation) ont joué un rôle crucial dans le processus de prise de décision, en garantissant à l'équipe Hannover Re la protection de ses données.

### Réduction des coûts et de la complexité

Hannover Re s'est débarrassé des outils ETL chers et lents, car MarkLogic ingère les données « telles quelles ». Grâce à la technologie d'indexage universel, l'équipe a pu réaliser tous ses objectifs de conception avec un seul produit plate-forme incluant une base de données et des fonctions de recherche et de programmation efficaces.

### Disponibilité et évolutivité des informations en temps réel

L'entrepôt de données analytiques centralisé a été conçu pour héberger des données provenant d'un grand nombre d'installations autour du globe. En raison des contraintes légales et de l'importance des analyses (séries chronologiques médicales), le système devrait collecter environ 10 années de données à partir de chacune de ces installations client. Globalement, le système est conçu et développé pour traiter efficacement des millions de documents en temps réel.

hannover re®





# THE WORLD'S BEST DATABASE FOR INTEGRATING DATA FROM SILOS



## STIWA AUGMENTE LE RENDEMENT DE PRODUCTION DE SES MACHINES D'AUTOMATISATION

Le Groupe STIWA, un des leaders de l'automatisation haute performance, fournit des systèmes automatisés d'emballage, de transport, d'assemblage et de fabrication. Les ingénieurs STIWA utilisent MATLAB® pour analyser de gros volumes de données de production collectés par le système de gestion de leur atelier (AMS ZPoint-CI), en temps réel. Cette approche permet d'optimiser le calcul des trajectoires des machines et de réduire la durée des cycles des systèmes de production (bras robotisés), tout en augmentant leur rendement.

Toutes les 8 secondes, un système standard STIWA collecte plus de 9 méga-octets de mesures de production brutes. Ces données doivent être filtrées et traitées pour identifier les tolérances appropriées, calculer les pistes d'optimisation de temps pour les robots ou les systèmes de transfert flexibles ou pour modifier le processus. Du fait de la complexité croissante des algorithmes nécessaires à leurs analyses, le besoin de s'équiper d'un environnement de développement comprenant des fonctions intégrées pour l'analyse des données est devenu évident, afin que les ingénieurs puissent se concentrer sur la résolution des problèmes et non sur des détails d'implémentation de bas niveau. Cet environnement devait permettre d'intégrer facilement les algorithmes au sein de leur système de production existant tout en tirant profit de l'infrastructure matérielle multi-cœurs disponible.

En utilisant l'environnement MATLAB, les ingénieurs ont réussi à automatiser l'analyse des données, ainsi que la planification des trajectoires des bras robotisés. Leurs algorithmes d'optimisation de trajectoire tirent profit de l'Optimization Toolbox™, les calculs sont accélérés avec la Parallel Computing Toolbox™ et les techniques implicites de 'vectorisation' de MATLAB. MATLAB Profiler leur permet d'identifier les fonctions les plus chronophages, puis de les valider via des frameworks de tests unitaires. Le passage en production se fait via MATLAB Compiler™ et Compiler SDK™ pour transformer leurs algorithmes en composants .NET qui, après une 2ème phase de test, sont intégrés au système de gestion de l'atelier. Toutes les 10 secondes ces algorithmes sont invoqués pour analyser les données collectées et calibrer les machines. Résultat :

- « MATLAB nous a permis de réduire de 30% la durée totale du cycle pour les processus de machinerie, résultant en une augmentation significative du rendement de la production », selon Alexander Meisinger CEO.
- « Avec MATLAB Compiler SDK pour intégrer nos algorithmes MATLAB dans un système de production .NET, nous pouvons facilement déployer l'algorithme sur plusieurs machines » conclut Robert Schoßleitner, Responsable de l'unité du logiciel de fabrication chez STIWA



## MONDI IMPLÈMENTE UN SYSTÈME DE MAINTENANCE PRÉDICTIVE DE SES PROCESSUS DE FABRICATION GRÂCE AU MACHINE LEARNING

Mondi Gronau est un leader mondial de la fabrication d'emballages et de produits à base de papier. Les 900 employés de son usine de production de plastique exploitent une soixantaine de machines d'extrusion, impression, collage et de bobinage, 24h/24, 365 jours par an. Mondi a développé un outil interne de suivi de fonctionnement et maintenance prédictive basée sur des statistiques avancées et le Machine Learning, le but étant de permettre aux employés de mettre en place des actions correctives et prévenir de sérieux problèmes.

Les machines utilisées par Mondi sont larges et complexes et sont chacune contrôlées par plusieurs contrôleurs logiques programmables, qui consignent des mesures telles que la température, pression et autres paramètres de performance provenant des capteurs. Chaque machine enregistre 300 à 400 valeurs par minute pour chaque paramètre, générant 7 Giga-octets de données quotidiennement. Les ingénieurs devaient développer une application permettant de présenter les résultats de ce type d'analyse, de manière claire et immédiate, aux opérateurs des machines. Et à des fins d'utilisation en continu en environnement de production, il fallait la structurer de manière adéquate, le tout avec peu d'expérience en analyse statistique et Machine Learning.

Avec l'aide du Consulting MathWorks, Mondi a développé une application MATLAB pour effectuer des requêtes vers leur base de données volumineuse, nettoyer les signaux et reporter graphiquement les résultats. L'application a ensuite été améliorée, en y intégrant des fonctionnalités de contrôle de processus statistique (SPC) alertant les opérateurs d'éventuelles valeurs hors normes.

En utilisant la Statistics & Machine Learning Toolbox™ et la Neural Network Toolbox™, et plusieurs algorithmes de Machine Learning (réseaux de neurones, k-means, les arbres de décision, SVM...), l'équipe a pu tester rapidement différents modèles de prédiction, puis incorporer le meilleur à l'application. Ces prédictions permettent aux opérateurs de recevoir des alertes sur de potentielles pannes. Mondi a ensuite utilisé MATLAB Compiler™ pour déployer ces applications en production dans l'usine. Résultat :

- « Plus de 50000€ économisés chaque année grâce à la maintenance prédictive avec MATLAB » déclare Dr. Kohlert, Head of information management and process automation chez Mondi
- « Nous avons eu les premiers tests en deux mois et un prototype opérationnel en six mois. Le code MATLAB est facile à comprendre, de sorte que nous pouvons faire des changements rapidement si besoin »





### OÙ EN EST, D'APRÈS VOUS, L'ADOPTION DES TECHNOLOGIES DITES 'BIG DATA' ?

On croise 2 profils d'entreprises : celles connues comme 'early adopters' bien avancées dans les projets 'Big Data', et les 'early majority', plus prudents. Ces derniers, sont plutôt en attente de résultats prouvés dans leur domaine avant de se lancer, surveillent de près les évolutions, et investissent prudemment dans des PoC. De nombreuses entreprises ont conscience de pouvoir accéder à des données à partir de sources multiples et utiliser des outils sophistiqués d'analyse prédictive, comme le machine learning, pour prendre rapidement des décisions plus efficaces. Nous le constatons dans presque toutes les disciplines de l'ingénierie, que ce soit dans le secteur médical ou l'aéronautique. Dans le secteur automobile, les entreprises acquièrent des données sur les passagers et les véhicules pour que leurs ingénieurs puissent optimiser la consommation de carburant de véhicules hybrides. Ceci est révélateur de la multiplicité des sources dont sont issues les données collectées, et où la construction de modèles complexes permet de mieux comprendre comment un système fonctionne dans le monde réel. L'utilisation de ces modèles de conception de système est un gain de temps et d'argent car il influe directement sur l'optimisation des performances opérationnelles. Cependant, les résultats sont encore timides. Des études, comme celle menée par Cap Gemini et Informatica qui révèle que 'moins d'un tiers des projets sont rentables', montrent que les entreprises ont devant elles une grande marge de progression. L'accompagnement des clients est donc clé pour inverser la tendance.

### QUELS SONT LES ÉLÉMENTS BLOQUANTS OU DÉFIS TECHNIQUES POUVANT EXPLIQUER CES FAIBLES RÉSULTATS ?

Nous pouvons scinder les défis en deux catégories principales. Tout d'abord, les défis d'ordre organisationnels : difficultés de communication entre différentes équipes, business vs analytique. S'ajoute à cela l'aspect gouvernance pour éviter les recouvrements de tâches entre DSI et départements analytiques/digitaux [généralement créés pour prendre le relais sur ces problématiques]. D'ailleurs, l'étude Cap Gemini-Informatica stipule qu'un projet aurait deux fois plus de chances d'aboutir si orchestré par une unité 'digitale' (vs DSI). D'autre part, d'un point de vue technique : la qualité de la donnée elle-même ('clean data') est capitale, mais surtout la 'mise en production' ou passage au monde réel. Cela se traduit par la difficulté à intégrer ces nouvelles couches applicatives développées (ingestion de donnée brute en temps réel, algorithme de prédiction, scoring, ...) aux environnements existants. Un autre type de défi tient à la préférence des entreprises d'avoir leurs propres experts du domaine capables d'effectuer les tâches Data Analytics requises - comme le prétraitement des données, le développement de modèles prédictifs, etc. - plutôt que d'embaucher des data scientists qui ont certes les compétences techniques mais pas l'expertise du domaine. Et c'est ce que nous permettons, notamment grâce à l'ensemble du support technique disponible à travers la formation, les séminaires et notre communauté d'utilisateurs en ligne MATLAB Central.

### EN TERMES DE TYPE D'OUTILLAGE, OÙ SE POSITIONNENT VOS SOLUTIONS ? (« Plateforme vs. Code »)

Les solutions se résumant exclusivement soit au code soit à l'interface utilisateur ont toutes deux leurs limites. MATLAB trouve l'équilibre qui permet aux experts du domaine de développer le Data Analytics et de travailler avec des Big Data. Nous proposons de nombreuses Apps qui offrent une expérience interactive pour accéder, visualiser, prétraiter, et modéliser des données - qu'elles soient numériques, de signal, ou encore provenant d'images ou de vidéos. Un exemple est la Classification Learner App, qui vous permet d'explorer vos données, sélectionner les fonctionnalités, spécifier la validation de schémas, choisir parmi un certain nombre d'algorithmes de machine learning, d'entraîner des modèles, et d'évaluer les résultats, le tout de manière interactive. Cette App MATLAB et d'autres génèrent également du code MATLAB, ce qui est important puisque la programmation haut niveau permet l'automatisation de tâches répétitives qui peuvent être fastidieuses avec une solution centrée sur l'interface utilisateur.

### COMMENT VOTRE OFFRE PERMET-ELLE D'ADRESSER LES DIFFÉRENTS PROBLÈMES CITÉS ? ET COMMENT ÉVOLUE-T-ELLE ?

Pour ce qui est de la qualité de la donnée, dans le monde de l'ingénierie et des objets connectés, une des solutions consiste aujourd'hui à embarquer des algorithmes de traitement à la source (le capteur lui-même), cela réduit considérablement la volumétrie de données transmises et coûts associés. Nos outils de génération automatique de code embarqué sont très utiles dans ce contexte. Quant à la mise en production, rapide, si je rebondis sur le constat établi par rapport au pilotage des projets par les DSI : ces derniers utilisant des langages et environnements différents des équipes analytiques (experts métier, data scientists...), la communication est implicitement rendue difficile. A cela, nos solutions permettent de réduire ce fossé entre métiers et DSI en automatisant le déploiement et la mise à l'échelle d'analytique MATLAB vers tout type d'environnement. Quant à l'évolution de notre offre, il est vrai que dans ce monde-là les versions applicatives changent très vite. Il n'en demeure pas moins que pouvoir en tirer profit n'est pas à la portée de tous. Notre approche de développement consiste à proposer, à un rythme réfléchi et prudent, des API facilitant l'interfaçage avec ces différentes technologies et une implémentation simplifiée des nouveaux frameworks tels que MapReduce et Spark. Car en tant qu'éditeur de logiciel avec des exigences qualité strictes, il serait coûteux et risqué de livrer des API sur des versions non-stables au prétexte de vouloir à tout prix supporter toutes les nouveautés technologiques du marché.

Référence : «The Big Data Payoff: Turning Big Data into Business Value» © Capgemini - Informatica 2016



## EN UNE PHRASE, MATLO C'EST QUOI ?

Matlo est une start-up spécialisée dans la business intelligence qui propose aux entreprises une plateforme d'analyse de données à partir de la data visualisation.

## POUVEZ-VOUS NOUS PRÉSENTER VOTRE SOCIÉTÉ ?

Nous avons créé Atelier Iceberg, une des premières agences de data design interactif en France en 2013 avec Thomas Dupeyrat, designer comme moi. Trois ans après, nous transformons l'agence en éditeur de logiciels de business intelligence pour devenir Matlo. Après une levée de fonds clôturée en un mois, nous sommes aujourd'hui 10 à bord pour continuer à accompagner les entreprises dans leurs problématiques d'analyse et de compréhension des données.

## QUEL REGARD PORTEZ-VOUS SUR VOTRE MARCHÉ ?

Le marché mondial de la business intelligence est en forte croissance depuis plusieurs années. Nous nous positionnons sur le segment de la data discovery très concentré par de nombreux éditeurs de logiciels nord-américains. Le nombre d'entreprises en France qui se lancent dans l'analyse visuelle de données reste encore restreint. Etant donné les retours que nos clients nous font aujourd'hui, nous sommes très optimistes quant à la poursuite de la croissance de la data discovery en France comme à l'étranger.

Bien qu'historique, la business intelligence apparaît encore aujourd'hui complexe aux yeux des entreprises. Notre rôle est de simplifier les usages liés à ces technologies et de continuer à démocratiser les fonctions Big Data et d'analytique. Le transfert récent des fonctions d'analyse de données des Directions informatiques vers les Directions métiers a fait naître des besoins nouveaux : autonomie, simplicité d'usage et collaboration. Aujourd'hui, ce sont les métiers eux-mêmes qui explorent, manipulent et restituent leurs données afin d'améliorer la performance de leurs activités et d'optimiser leur prise de décision. Et c'est notre conviction : nous voulons permettre à tous les métiers d'analyser eux-mêmes et rapidement leurs données pour mieux les comprendre.

## QUELLE EST LA DIFFÉRENCE DE MATLO PAR RAPPORT AUX ACTEURS DÉJÀ PRÉSENTS SUR LE MARCHÉ ?

Nous proposons un socle commun de fonctionnalités par rapport aux autres outils de BI et faisons en plus un double pari. D'abord, nous voulons aider nos clients à mieux déchiffrer leurs données avant même de commencer à créer leur tableau de bord. Pour cela nous avons créé un moteur de suggestions qui repère dans le fichier du client les données les plus importantes et les plus corrélées pour lui proposer de commencer son analyse par ces données. Nous voulons aussi que tout le monde puisse utiliser Matlo et construire soi-même son tableau de bord. Nous cherchons donc à créer l'environnement de travail le plus ergonomique et agréable possible pour permettre au client de gagner du temps dans l'utilisation de Matlo et de prendre du plaisir à explorer ses données. Nos designers font tout pour éviter l'apparence austère des outils de business intelligence historiques. Ils proposent une interface épurée et testent en permanence des améliorations ergonomiques pour nous assurer que même sans compétence technique nos clients peuvent comprendre facilement leurs données.

## A QUI EST DESTINÉ MATLO ?

Nous avons parmi nos clients des PME et des grands comptes dont certains sont cotés au CAC 40. Plutôt que de cibler des secteurs d'activités, nous nous adressons d'abord à des métiers comme le marketing, les études, les RH, la veille, la stratégie, la R&D et plus généralement à tous ceux qui ont besoin d'extraire des informations de qualité de leurs données.

## QUELLES SONT LES GRANDES ÉTAPES DE DÉVELOPPEMENT DE MATLO ?

Après trois ans de prestations de design interactif et de conseil en data visualisation nous avons validé une méthodologie qui nous permet aujourd'hui d'être sereins sur le logiciel que nous proposons à nos clients. Pour valider les retours et évaluer les besoins exprimés par nos clients ces trois dernières années, nous avons réalisé en 2015 une trentaine d'entretiens auprès de grands comptes, d'agence de consulting, de marketing, de veille et des institutions.

Après dix mois de développement informatique, nous avons lancé Matlo en version Alpha auprès d'entreprises comme L'Oréal, BVA, Capgemini Consulting, Fabernovel, The Links ou Ylios pour ajuster le produit selon leurs retours et s'assurer que ce que nous faisons est véritablement utile. Matlo est alors disponible en version Bêta jusqu'à ce que nous propositions notre offre cloud enrichie de nouvelles fonctionnalités dès janvier 2017. Cette approche « prestation - enquête - version Alpha - version Bêta » nous assure en permanence d'être au plus près des besoins de nos clients.





EXPLOREZ  
VOS DONNÉES  
POUR MIEUX  
DÉCIDER

Découvrez la nouvelle  
plateforme d'analyse  
de données sur :

[www.matlo.com](http://www.matlo.com)



## MICROPOLE

David FALA

Directeur Practice Big Data et Data Intelligence

+33(0)1 74 18 74 18

info@micropole.com

91-95 - rue Carnot 92300 LEVALLOIS-PERRET France



www.micropole.com

**MICROPOLE**  
Driving Distinction


## LA DATA INTELLIGENCE, LA NOUVELLE VALEUR DU BIG DATA

Pour faire face à la suprématie que les champions de la nouvelle économie ont souvent bâtie sur l'exploitation de données massives, beaucoup d'entreprises sont en quête du business model de rupture qui leur permettra de trouver leur salut. Le véritable challenge auquel doivent faire face ces entreprises est clair : innover et se réinventer pour rester compétitives.

Cependant, les entreprises sont confrontées à un vrai dilemme face à l'avènement des technologies Big Data qui promettent de rendre intelligibles et intelligents tous types de contenus : faut-il comme le prétendent certains, faire table rase du passé et des dizaines d'années de reporting, de statistiques, de business intelligence et de datawarehouse, pour les remplacer par un SI Hadoop-centrique, dopé au machine learning et nourri sans discernement de n'importe quelles données ?

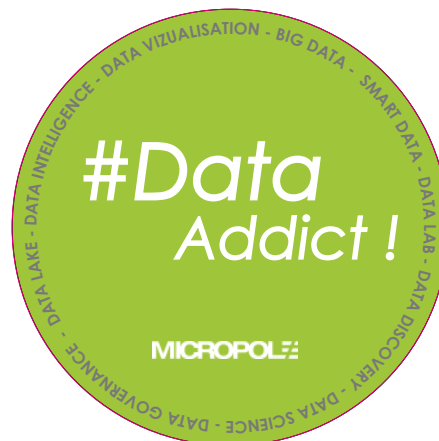
La réponse des équipes Big Data de Micropole est sans appel : la valeur n'est pas dans l'accumulation de technologies et de données mais dans le traitement raisonné, légal et éthique de toutes les données utiles. À l'ère de la gouvernance nouvelle entre les métiers et l'informatique, qui rend l'entreprise agile dans la valorisation de son patrimoine, et alors même que la technologie ambiante rend presque tout envisageable, la clé de la performance réside dans une démarche globale de DATA INTELLIGENCE, et il faut que chaque entreprise en pose les fondations.

Pour ce faire, Micropole :

- Sensibilise et accompagne ses clients sur la mise en œuvre de cette nouvelle gouvernance (Data Office, outils, méthodes, utilisation des techniques collaboratives).

- Aide à construire les organisations de test-and-learn dotées de moyens techniques, comme les Datalabs, pour permettre la découverte de nouvelle valeur au traitement de la donnée.
- Encourage et supporte les intégrations de Datalake dans les infrastructures.
- Forme et organise les nouvelles équipes de DataScience en fonction des besoins réels et des compétences évaluées.
- Assiste à la découverte et l'incorporation dans le SI de données très volumineuses internes (logs, IoT, ...) et externes (opendata, clouds divers, fournisseurs de data services divers ...).
- Étend la performance souvent critiquée de la BI en mettant en œuvre tous les nouveaux concepts d'Analytics nécessaires à une organisation en fonction de sa maturité et de ses objectifs.

ET VOUS, OÙ EN ÊTES-VOUS DANS VOTRE TRANSFORMATION DATA INTELLIGENTE ? PARLEZ-EN A NOS #DATA ADDICTS !



## Quelques-unes des missions réalisées par nos #Data Addicts

## SERVICE PUBLIC

**Problématique client** : éduquer et aguerrir les équipes statistiques d'analyse et de prospective économique à l'utilisation en mode DataScience de logiciels prédictifs et de machine learning en environnement BigData et DataViz.

**Notre réponse** : sensibilisation et formations théoriques et techniques. Organisation de cellule Datalab et mise en œuvre des moyens prédictifs autour d'Hadoop.

## INDUSTRIE

**Problématique client** : gouverner le patrimoine informationnel industriel et commercial et dépasser les clivages d'un SI fortement distribué techniquement et géographiquement, pour en tirer plus de valeur.

**Notre réponse** : mise en œuvre d'une organisation de gouvernance des usages de la donnée, d'une convergence des sources multiples vers un Datalake Hadoop, avec mise en œuvre d'un iPaaS entre le SI on-premise et les différents clouds.

## ASSURANCE

**Problématique client** : création d'un Observatoire des données statistiques et financières pour le secteur.

**Notre réponse** : mise en place du socle technique BigData / DataScience. Datalake d'accueil des données internes et OpenData, anticipation de collecte de données IoT.

## LUXE

**Problématique client** : enrichir la connaissance client multi-produits pour affiner l'efficacité des campagnes marketing personnalisées.

**Notre réponse** : mise en place d'une logique stricte de référentiel client (MDM) et association en 360° toutes les données internes et externes avec une démarche Data Quality étendue aux métiers (DataStewardship).

## TRANSPORT LOGISTIQUE

**Problématique client** : accroître la fidélisation clients, la performance multimodale, la qualité de service et la traçabilité. Mais aussi préparer la monétisation par la Data as a Service et enrichir les données et les moyens de planification stratégique.

**Notre réponse** : machine learning et collecte systématique des données (IoT renforcé), gouvernance des données et encouragement au « fail fast » pour innover sans retenue. Exploitation systématique de streaming et de Hadoop.

## BANQUE

**Problématique client** : lutte contre la fraude, meilleure connaissance des clients et de leur parcours multicanal, réduction des coûts de data management.

**Notre réponse** : définition et mise en place d'un datalake sous Hadoop pour recevoir les données depuis les mainframes. Gestion des contenus (dématérialisation) et analyse des sentiments clients sur les réseaux sociaux.

# Christian Blanc

Directeur des Systèmes d'Information

## INTERVIEW



« Nous sommes poussés par les métiers à fournir des solutions en 3-4 mois, là où on met d'habitude 16 - 20 mois. Alors, pour garantir la mise à disposition de données propres cohérentes et de qualité, on s'appuie sur des notions de DataLake et de traitement au sens BigData de la donnée. »

### CHRISTIAN BLANC, VOUS ÊTES LE DIRECTEUR DES SYSTÈMES D'INFORMATION DE BNP PARIBAS WEALTH MANAGEMENT, UNE DES PREMIÈRES BANQUES DE GESTION PRIVÉE DU MARCHÉ. POUVEZ-VOUS NOUS PARLER DE VOS ENJEUX ET PROJETS BIG DATA ?

Notre système d'information est complexe et s'est construit au cours du temps par couches successives. Notre maîtrise des données doit répondre à deux exigences principales : le contrôle financier et la conformité réglementaire d'une part ; la performance analytique et le time-to-market pour nos métiers, d'autre part. Dans le cadre du plan stratégique de transformation du groupe mené par la Direction de BNP Paribas, nous avons entamé avec Micropole un indispensable chantier de cohérence autour de nos référentiels d'entreprise, et de notre référentiel clients en particulier. Lorsque la vague de la mode Big Data a commencé à déferler sur les métiers du marketing, les directions fonctionnelles ont immédiatement poussé à investir massivement sur ces nouvelles technologies, avec un objectif : connaître toujours mieux nos clients. Passés les premiers POC technologiques avec notre partenaire Micropole et la DSI, les métiers ont tout d'abord peiné à déterminer des pistes de développement concrètes. Faute d'urgence, nous avons décidé de nous consacrer au sujet DATA en global afin de tirer parti du périmètre de données déjà riche et hétérogène du Customer Journey Analysis. Même si cela ne représentait pas des volumes phénoménaux, ce projet nous a permis d'envisager une analyse systématique et sémantique de tous types de données grâce justement aux méthodes et techniques issues de la recherche Big Data.

Pour ce faire, le concept de Data Lake proposé par Micropole, permettait d'homogénéiser les données et d'offrir un point de concentration pour toutes les données et contenus « chauds et froids » du périmètre client. Les techniques analytiques du monde Big Data ont ainsi permis d'envisager d'autres moyens pour tirer plus de valeur des données que la classique BI que nous utilisons depuis 30 ans.

### VOUS AVEZ DONC DÉCIDÉ DE RATIONALISER LES PROJETS EN MUTUALISANT LES CAPTURES DES DIFFÉRENTS SYSTÈMES CIBLÉS AU SEIN D'UN GISEMENT DE DONNÉES UNIQUE. COMMENT VOUS Y PRENEZ-VOUS ?

Il y a 3 niveaux de réponse à cette question : la technique, la méthodologie, l'organisation. Nous avons ainsi classiquement utilisé la méthode du Data Lake. L'opération a été technique mais pas si compliquée lorsque l'on s'appuie sur les bonnes compétences d'équipes bien organisées. La méthodologie employée a été guidée par deux principes que sont : Agilité et Gouvernance.

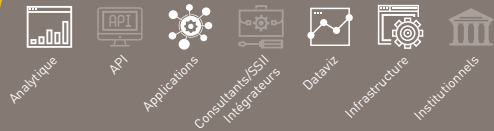
Le Data Lake a été créé sur un périmètre fonctionnel qui couvre les besoins actuels autour du Customer Journey, et prévu pour augmenter au fur et à mesure des nouveaux périmètres à couvrir. L'exigence d'agilité demande de procéder par itérations continues.

Ce périmètre, actuel et futur, a été défini avec les métiers sur la base des sources de données qu'il importe de concentrer pour répondre à leur besoin. Tout ce qui aujourd'hui entre dans le Data Lake est documenté techniquement. Les métiers sont donc sollicités en permanence pour documenter les données et leurs usages. C'est un véritable travail de gouvernance, crucial pour la capitalisation autour du « patrimoine des données ». Enfin, au niveau de l'organisation, le Chief Data Officer et le Data Custodian ont dirigé les travaux avec leurs correspondants, l'un du côté fonctionnel et l'autre en contrepartie technique.

Au niveau opérationnel, l'organisation d'ateliers de valorisation rapide des données, dans des modèles collaboratifs de type « Pizza Team », est très efficace dès lors que les données nécessaires sont immédiatement disponibles. On fait alors évoluer le rapport classique MOA/MOE, avec un modèle plus adapté à l'implication multipartite, -business, compliance, juridique, informatique, ... -, et au résultat rapide, qui remet en question fortement nos méthodes de travail classiques. Avoir sous la main les données, les outils analytiques et les bonnes ressources internes et externes, nous permet d'atteindre en permanence nos deux objectifs clés.

**Zahir Abdelouhab**  
Regional Director

+33 (0)1 82 881 666  
zahir.abdelouhab@mongodb.com  
171 bis Avenue Charles de Gaulle  
92210 Neuilly-sur-Seine



www.mongodb.com



## RETHINKING THE CUSTOMER EXPERIENCE AT METLIFE

### FROM STALLED TO SUCCESS IN THREE MONTHS

MetLife wanted to harness Big Data to create a **360-degree** view of its customers so it could know and talk to each of its more than 100 million clients as individuals. But the Fortune 50 company had already spent many years trying unsuccessfully to develop this kind of centralized system using relational databases.

**Which is why the 145-year old insurer in 2013 turned to MongoDB. Using MongoDB's technology over just 2 weeks, MetLife created a working prototype of a new system that pulled together every single relevant piece of customer information about each client.**

Three short months later, the finished version of this new system, called the 'MetLife Wall,' was in production across MetLife's call centers. **The Wall** collects vast amounts of structured and unstructured information from MetLife's more than 70 different administrative systems. After many years of trying, MetLife solved one of the biggest data challenges dogging companies today. All by using MongoDB's innovative approach for organizing massive amounts of data.

### INSIDE THE METLIFE WALL

Today, when a customer calls MetLife to ask about a claim, add a new baby to a policy, or dig into coverage details, customer representatives use the Wall to pull up every bit of information they need in seconds – name, address, policies, and life events.

Using a touchscreen and a design based on how Facebook dishes up information, The Wall is instantly familiar to MetLife's call center operators. Which means customer reps can quickly and efficiently answer questions, handle claims, suggest new services, or offer promotions, while slashing wait and call times. MetLife now understands and serves each customer individually.

### POWER OF THE FLEXIBLE DATA MODEL

Like most companies, MetLife has scores of data systems created or acquired over the years. MetLife's systems contain a huge array of structured and unstructured data, including policy and customer information and transactional history about everything from claims to payments. Few are connected and many are on mainframes with cumbersome interfaces. Ripping out its administrative systems and replacing them with one unified system wasn't an option for MetLife. So the company had tried over the years to use relational databases, which require a common schema and strict mapping of data sources. Adding each new system was an expensive and time consuming process of changing schemas, and extracting, cleansing, and matching data – one that MetLife never won.

### METLIFE'S CHALLENGE: DATA VARIETY

MetLife's 70+ administrative systems contain a massive variety of structured and unstructured data, of two main types. One includes 50 million policies and 118 million customers. Another represents transactional history about payments and claims, and includes about 190 million documents. MetLife needed a way to pull this all together in a single view.

### RELATIONAL DATABASE DIDN'T WORK

A relational database resembles many Excel spreadsheets. It has a highly structured table organization (its schema). When it comes to customers, for example, you know different things about each one. You may have just the name and email of some, while for others you also know their phone number and different shipping addresses. To make this fit into a spreadsheet, you need to create lots of columns, many of which will be empty. This database becomes unwieldy and difficult to manage.

## MONGODB'S DOCUMENT APPROACH WORKED

MongoDB stores information like a series of Word documents. A set of data is stored in a document that has its own schema. When you add a field to a given dataset, you can do so without having to add that field to all of the other documents. For instance, when it comes to managing customer data, you'd use a document for each individual. Everything you know about that person is stored in the document. Some documents have just a few fields, while others contain a lot of information. Adding new information about one customer doesn't require updating all the other documents.

What makes MongoDB different is its flexible data model. MongoDB looks at data more naturally, making it easy to evolve schemas in real time. If relational databases are like Excel spreadsheets – where data is organized into sheets but where you add a column every time you add a field, creating a structured but unwieldy project – MongoDB is a series of Word documents. Each entry is a document that can have its own schema.

### FLEXIBLE, SCALABLE, USER FRIENDLY

**MongoDB's horizontal scaling via automatic sharding provides reliable partitioning of massive amounts of data across multiple servers. And it's flexible, allowing organizations to leverage multiple data centers and multi-temperature storage techniques.**

**Just as crucial for productivity and agile application development is the ease of use MongoDB provides developers. Developers can interact with the database in the same programming language they use to write the application, whether Java, Ruby, Python, or something else. Which means they can focus on building apps instead of wrestling with ORMs.**

**And MongoDB provides a variety of rich features, such as integrated search, geospatial, and native analytics, that don't exist in a traditional database. Giving companies the right resources they need to get projects done quickly.**

### A CATALYST FOR INNOVATION

Working with MongoDB led not just to The Wall's success, but a wave of innovation inside MetLife. Harnessing Big Data is letting the insurance giant do things it never could do before.

### NOW, METLIFE IS:

- Rolling out The Wall internationally
- Developing a version of The Wall for its sales force
- Creating a real-time analytical system that predicts customer attrition rates, prompting customer reps to offer alternative products or promotions
- Considering adding social data and data from mobile apps into The Wall to gain an even better understanding of customers, something that's easy to do with MongoDB.

With MongoDB, MetLife joined the vanguard of the Big Data revolution. By quickly and successfully cracking one of the biggest problems it faced, MetLife changed the expectations of what is possible.



# Optez pour une base de données nouvelle génération.

Soyez prêt à relever les plus  
grands défis de votre entreprise.

**Cédric FAUVET**

Business Développement France

+33 (0) 8 05 08 03 44

ventes@neotechnology.com

6, rue Duret - 75116 Paris



www.neo4j.com



## COMMENT NEO4J A AIDE A TRAITER LES DONNEES « PANAMA PAPERS » LA PLUS GROSSE FUITE DE DONNEES FINANCIERES DE L'HISTOIRE

Le Consortium International de Journalistes d'Investigation (ICIJ) a utilisé la base de données de graphes Neo4j pour analyser ce qui est considéré comme la plus grosse fuite de données financière connue dans l'Histoire.

Récemment révélée par le Consortium International de Journalistes d'Investigation (ICIJ) l'affaire « Panama Papers » est le scandale qui éclabousse de nombreuses élites mondiales et a mis à jour le système des activités offshore dans les paradis fiscaux.

Les données qui ont fuité décrivent les opérations internes du panaméen Mossack Fonseca, l'un des cabinets les plus réputés au monde en matière de création de sociétés offshore. Les 2,6 Téraoctets de données qui composent le dossier ont été obtenues par le quotidien allemand Süddeutsche Zeitung et partagées avec l'ICIJ basé à Washington, qui est un réseau d'équipes d'investigations indépendantes internationales qui collabore avec plus d'une centaine de médias de tous les pays.

La pierre angulaire des nombreux scoops révélés par l'ICIJ, à l'image de l'affaire Swiss Leaks, n'est autre que Neo4j. C'est en fait cette technologie de base de données de graphes, utilisée avec la plateforme de visualisation Linkurious, qui a été utilisée pour organiser, orchestrer et rendre intelligibles ces données hautement interconnectées.

### Un volume de données connectées sans précédent

Selon Mar Cabra, Data and Research Unit Editor de l'ICIJ, l'utilisation de Neo4j était la seule solution viable et disponible pour répondre à ces besoins. « C'est un outil de recherche révolutionnaire qui transforme le processus d'investigation des journalistes car dans ce cas, les relations sont extrêmement importantes puisque ce sont elles qui nous montrent où sont les criminels, qui travaille avec qui, et bien plus encore. Comprendre ces relations à une très grande échelle est précisément là où les bases de données graphes excellent », explique-t-elle.

« Avec 11,5 millions de documents à traiter, bien plus que n'importe quelle fuite de données précédente, nous avons donc besoin d'une technologie qui puisse gérer un tel volume rapidement, facilement et de manière efficace », ajoute-t-elle.

## Ouvrir l'accès aux données Panama Papers pour plus de transparence

Après les différentes révélations faites par l'ICIJ et ayant eu de nombreuses répercussions tant pour des célébrités que des personnalités publiques, le consortium a mis à disposition les données issues des Panama Papers via une application basée sur Neo4j. Les données divulguées contiennent notamment des informations sur plus de 200 000 sociétés offshores étudiées dans le cadre de l'enquête.

Avec la mise à disposition de cette application, tout le monde peut désormais accéder d'une part aux données disponibles et visualiser les milliers de structures offshores, d'autre part aux enregistrements internes de Mossack Fonseca dont les noms des vrais propriétaires des dites structures. Propulsée par Neo4j, la base de données interactive comprend également des informations sur plus de 100 000 structures offshores qui figuraient déjà dans des documents dévoilés par l'ICIJ en 2013.

La démocratisation de technologies capables de donner du sens à de telles masses de données interconnectées est une composante inévitabile dans une société libre et ouverte, or les bases de données de graphes sont les seules solutions viables lorsqu'il s'agit de donner du sens à des téraoctets de données connectées entre elles. Elles sont aussi essentielles pour les agences internationales, les gouvernements, les services financiers et les spécialistes de la sécurité à la recherche de la découverte de la vérité.

En cela, Neo4j a joué un rôle très important, que ce soit dans les Swiss Leaks ou dans les Panama Papers, et continuera de le faire à l'avenir.



De gauche à droite  
Stéphane Marlin  
Benoît Simard  
Cédric Fauvet



## INTERVIEW

**QUEL EST LE LIEN ENTRE GRAPHES ET BIG DATA ?**

Les entreprises ont de plus en plus de données à traiter. Elles oublient souvent que la valeur de ces quantités massive de données réside aussi dans les relations qui les unissent. Hors les outils d'analyse des données SQL ou No SQL classiques ne peuvent traiter les relations dans le volume avec performance et simplicité, ce qui explique le succès de la base de données de graphes Neo4j.

**VOUS AURIEZ UN EXEMPLE CONCRET A NOUS DONNER ?**

Bien-sûr. Si l'on prend l'exemple d'HADOOP. Il permet de stocker des logs en grande quantité. Remonter des problèmes présents dans les logs est effectivement très intéressant pour l'entreprise, l'étape suivante est de comprendre pourquoi ces problèmes sont apparus. Une base de données classique ne saura pas analyser la topologie d'un réseau électrique, par exemple, et ainsi ne pourra pas permettre d'identifier les risques de problèmes en cascades. Avec une base de graphes, on a le problème, la topologie du réseau, les interactions entre l'endroit où le problème est apparu et toutes les interdépendances entre cet endroit et les autres éléments. L'entreprise va ainsi pouvoir rapidement résoudre le problème et corriger ce qui doit l'être pour éviter un impact ailleurs.

**CELA SIGNIFIE-T-IL QUE SELON VOUS LE NOM 'BIG DATA' EST SUR-EVALUE ?**

Non, du tout. Simplement, je pense qu'aujourd'hui, nous abordons en quelques sortes un Big Data 2.0, dans lequel on découvre que les relations entre les données sont plus importantes et génèrent plus de valeur que les données seules. C'est un peu comme si l'on regardait le ciel la nuit, étoile par étoile. C'est intéressant, mais le ciel prend une autre dimension lorsqu'on comprend comment elles se positionnent les unes par rapport aux autres, permettant ainsi découvrir la Grande Ourse, l'étoile polaire... Mais sans un passage par le Big Data 2.0, nous n'en aurions pas pris conscience.

**EN SUIVANT VOTRE ANALOGIE SUR LE CIEL, APPARAÎT DEPUIS PEU LA NOTION DE 'DARK DATA', VOUS POUVEZ NOUS EN PARLER ?**

Oui, il s'agit d'une notion qui fait encore l'objet de discussions entre les analystes. Mais de façon générale, il s'agit des données potentiellement utiles qui pourraient être obtenues à partir de processus métier mais qui ne sont actuellement pas mises à profit. Et ces données posent problèmes, puisque le stockage a un coût et stocker des données non utilisées est un gaspillage de ressources. Si l'on tire le fil de cela, on peut estimer que les relations entre les données sont elles-mêmes des 'dark data' à l'heure actuelle. En effet, depuis longtemps, on a les outils pour trier et analyser les données. Toutefois, les entreprises ne disposaient d'aucun outil pour traiter les relations entre données. C'est ce qu'apportent les bases de graphes pour ces 'dark data' que sont les relations.

**QUI UTILISE AUJOURD'HUI LES GRAPHES ?**

Toutes les entreprises qui ont des projets avec des données interconnectées en volume. En effet, vouloir utiliser les graphes quand on n'a que très peu de données n'a pas de sens ; les bases de données traditionnelles font très bien le travail. En revanche, sur des données en nombre, les graphes prennent toute leur dimension et apportent une réelle aide à la décision. Aujourd'hui, par exemple, des entreprises comme LinkedIn, Transparency-One, Ebay, Walmart, Airbus, le groupe Adidas, Meetic ou IDVroom utilisent Neo4j.

**UN EXEMPLE MARQUANT A NOUS PRESENTER ?**

Il y en a de nombreux, puisque les graphes sont utilisés dans de nombreuses applications. Un exemple précis et que tout le monde a en tête est la révélation des Panama Papers. On se rappelle que le consortium international des journalistes d'investigation (ICIJ) a mis à jour 2.6 Téraoctets de données concernant des personnes ou sociétés qui faisaient de la fraude fiscale. C'est parce que l'ICIJ a utilisé Neo4j qu'elle a pu traiter et identifier les liaisons entre toutes ces données (personnes, banques, entreprises...). En effet, ces liaisons ont montré sans équivoque les différentes interactions et participations d'hommes dans des sociétés offshore, et le tout en un temps record.

**QUEL EST SELON VOUS L'AVENIR DES BASES DE DONNEES DE GRAPHES ?**

L'avenir des bases de données de graphes est tout simplement qu'elles soient utilisées dans tous les cas où une entreprise a à gérer des grosses quantités de données. D'ailleurs, cela correspond aux prévisions faites par Forrester Research dans une étude récente : cette année, la part de bases de données de graphes dans l'ensemble des bases de données mondiales devrait atteindre 25%. Et ce n'est que le début.

**POUR TERMINER, VOUS POURRIEZ NOUS PRESENTER VOTRE SOCIETE, NEO TECHNOLOGY, MEME SI NOUS AVONS BIEN COMPRIS QUE VOUS ETIEZ UN ACTEUR MAJEUR DES BASES DONNEES DE GRAPHES ?**

Neo Technology est le créateur de Neo4j, la base de données de graphes leader dans le monde, base de données de graphes native hautement évolutive qui permet de mettre à profit les relations entre les données. Cette technologie aide les entreprises à concevoir et mettre au point des applications intelligentes qui répondent aux besoins actuels et futurs des entreprises comme la détection de la fraude, les recommandations temps-réel, la gestion des données de référence, la sécurité réseau et les opérations des services informatiques. Avec ses origines Suédoises et son siège social établi à San Mateo en Californie, Neo Technology est présent en Europe et notamment en Suède, en Allemagne, en France et au Royaume-Uni.

**Stéphane DAUTEL**

Partner sponsor Big Data

+33(0)1 46 94 95 96

cbd@groupeonepoint.com

29 rue des Sablons 75116 Paris France



www.groupeonepoint.com

# onepoint.

## HISTORIQUE

Onepoint est devenu un acteur majeur de la transformation numérique. En tant qu'acteur international de taille intermédiaire, onepoint a su développer une culture agile nourrie d'influences canadiennes et de créativité française ainsi qu'une culture de l'engagement portée par des expertises et des moyens de production industriels en région, à Nantes et à Bordeaux. Onepoint ambitionne désormais de devenir un leader européen du Big Data.

## SAVOIR-FAIRE

Le métier de onepoint est autant de penser le changement que de le conduire avec les outils adaptés, pour des clients de tous secteurs. Notre offre intégrée nous permet d'apporter une réponse à chaque besoin et de nous mettre au service de sa transformation (innovation, digitalisation, agilité).

## CŒUR D'ACTIVITE

Concernant le Big Data, nos services et prestations s'organisent autour de 4 piliers :

**Smart Data** : conseil sur valorisation de la donnée métier dans des délais courts réclamant une mise en œuvre technologique limitée.

**Data Science** : notre connaissance des métiers et nos compétences algorithmiques associées à une utilisation de technologies de pointe, nous permet d'accompagner nos clients dans la valorisation d'un large ensemble de données internes et/ou externes.

**Optimisation du SI** : onepoint accompagne ses clients dans l'adoption des technologies Big Data dans le but de rationaliser et d'optimiser son SI d'un point de vue architecture et/ou coûts.

**Disruptive innovation** : notre expertise nous permet de projeter les innovations des technologies Big Data dans le contexte de nos clients et ainsi de leur offrir des solutions à des challenges passés ou à venir non adressables dans des technologies plus traditionnelles.

## SECTEURS CIBLES

Fort de l'expertise de plus de 1 800 collaborateurs, onepoint intervient dans tous les secteurs d'activité. Au cœur de notre organisation, nos communautés sont des espaces ouverts qui accueillent et font vivre nos expertises métiers, technologiques ou fonctionnelles. Nos communautés de services s'articulent aujourd'hui autour des secteurs Banque, Finance et Assurance, Services, Médias & Télécoms, Energie, Distribution, Transport, Secteur public et Lifescience.

## COUVERTURE GEOGRAPHIQUE

L'ouverture de onepoint à l'international a commencé dès 2003 en Amérique de Nord et plus particulièrement au Canada pour se poursuivre ensuite en Tunisie et en Chine (2007), au Royaume-Uni (2013) ainsi qu'au Luxembourg, en Belgique et aux Pays-Bas (2015).

## AVANTAGES CONCURRENTIELS

La force de onepoint est d'avoir conservé l'agilité d'une ETI innovante, sa capacité à proposer des solutions sur mesure à ses clients, grâce à des relations très fortes avec un écosystème de startups et de PME.

**Verticalité** : la diversité issue de nos communautés métiers et régions nous permet d'intervenir sur la conception et la réalisation de projets d'envergure dans le Big Data du conseil métier jusqu'à l'infrastructure.

**Partenariat** : dans le souci d'apporter de la visibilité à nos clients, onepoint a noué des partenariats stratégiques avec les meilleurs acteurs de l'écosystème NoSql ou de la fondation Apache : MapR, MongoDB et Elastic Search.

**Transmission** : onepoint a créé une académie dédiée de formation délivrant des certifications reconnues par nos partenaires. Afin que chacun puisse échanger et partager ses connaissances au sein de la communauté mais également en dehors, nous créons ou hébergeons des événements publics (Meetup, Hackathon, etc.).

## ONEPOINT REALISE DES MISSIONS BIG DATA POUR DE NOMBREUX CLIENTS.

### CI-DESSOUS, LE TEMOIGNAGE DE L'UN DE NOS CLIENTS : ORANGE

Orange a sollicité onepoint pour l'accompagner sur les premières phases d'alimentation de sa solution Big Data en raison de la qualité des précédentes missions conduites par onepoint, notamment sur l'exploitation et le maintien opérationnel de la plateforme.

J'ai apprécié chez onepoint la rapidité d'immersion dans un environnement et un écosystème 100% métier, à savoir les données du réseau mobile d'Orange et la capacité des consultants et de l'entreprise à travailler en mode agile (scrum), avec du développement itératif.

Ce développement a duré 6 mois. Cette expertise qui a été déployée permet également désormais à onepoint d'accompagner Orange sur son architecture Big Data et son dimensionnement, et très prochainement d'initier un volet Data Science (traitement de use cases, machine learning, corrélation de données).

**Témoignage de Othmane ALLAOUI – Data Scientist et Responsable de Programme Big Data Chez Orange - Octobre 2016**

## SUIVEZ-NOUS :





### « AMBITION : DEVENIR UN LEADER EUROPÉEN DU BIG DATA »

#### QUE SIGNIFIE POUR VOUS DEVENIR UN LEADER EUROPÉEN DU BIG DATA ?

Devenir un leader du Big Data signifie pour onepoint « être un acteur capable d'industrialiser les initiatives de ses clients ». Nous pensons en effet que le marché va largement se massifier et que les acteurs de référence seront ceux capables, comme onepoint d'accompagner leurs clients sur l'ensemble de la chaîne de valeur d'un projet : conseil métier, architecture, développement, production, infrastructure et ce durablement. Concrètement, cela signifie aussi que onepoint a pour ambition de faire grandir sa communauté Big Data de façon significative ces prochaines années afin d'accompagner la demande du marché.

#### COMMENT COMPTEZ-VOUS DEVENIR LE LEADER EUROPÉEN DU BIG DATA ?

Nous nous sommes dotés de partenaires technologiques de référence pour apporter une lisibilité à notre offre, nous pensons que trop d'attention et de temps est donné au benchmark technologique comparé à l'usage. Nous avons également acquis la conviction qu'une grande attention devait être mise sur l'opérabilité en particulier lors du passage en production des POCs, l'impact de la transformation étant souvent sous-estimé. Nos partenariats ont été pensés à cet effet. Enfin, nous souhaitons pallier la rareté des compétences, l'un des grands freins à l'adoption du Big Data. Pour ce faire, nous mettons en place une véritable académie de formation qui bénéficiera à l'ensemble des collaborateurs du groupe. Elle permettra de massifier nos compétences en certifiant nos consultants, en relation étroite avec nos partenaires éditeurs, présents sur le salon Big Data. Cette académie assure également la relation avec les écoles formant les meilleurs profils experts.

#### QUELS PARTENAIRES AVEZ-VOUS CHOISIS POUR ATTEINDRE CETTE AMBITION ?

Nous avons choisi d'industrialiser notre approche en optant pour une solution autour d'Hadoop opérable et performante : MapR. Nous avons également monté un partenariat avec MongoDB, largement utilisé par nos clients et répondant aux enjeux de time to market et d'agilité technologique dans un contexte transactionnel. Enfin nous nous sommes rapprochés d'ElasticSearch dont la suite technologie est une référence en particulier pour l'indexation chez un très grand nombre de nos clients. Nous développons également des compétences expertes avec des startups technologiques qui proposent des solutions sémantiques innovantes ou spécialisées dans le machine learning, deep learning, etc. Pour n'en citer qu'une : Semsoft.

#### POURQUOI PENSEZ-VOUS QUE LES EXPERTS BIG DATA CHOISIRONT ONEPOINT, PLUTÔT QUE D'AUTRES ENTREPRISES CONCURRENTES ?

Les experts choisissent onepoint pour notre capacité unique à leur offrir une variété de projets, de secteurs d'activité et de missions mais aussi une structure, un support et le temps libre nécessaire à consacrer aux projets qui leurs tiennent à cœur. L'organisation de l'entreprise basée sur les communautés et l'autonomie fait de onepoint une société dans laquelle l'initiative et l'innovation sont valorisées et encouragées. Le regroupement de profils experts favorise les échanges et les initiatives de sorte que chacun peut trouver sa place et contribuer à l'ambition du Groupe. Enfin, avec son académie, onepoint s'est doté d'un outil capable d'accompagner ses collaborateurs dans leur montée en compétences sur le domaine très vaste de la gestion de la donnée.

#### PORTES OUVERTES POUR DÉCOUVRIR LES ESPACES INNOVATION DE ONEPOINT. OUVERT À TOUS !

Prenez votre billet  
en scannant ce flashcode



# OUTSCALE

**Frédéric LEBLAN**

Account Manager

**+33(0)1 55 39 01 33**

sales@outscale.com

1, rue Royale - 319 Bureaux de la Colline  
92210 - Saint-Cloud France



[www.outscale.com](http://www.outscale.com)



## OUTSCALE, LA PREMIÈRE INFRASTRUCTURE CLOUD FRANÇAISE PRÉSENTE DANS LE MONDE

Fondée en France en 2010, partenaire stratégique de Dassault Systèmes, certifiée CMSP Advanced par Cisco Systems, 100% ICT par Intel et AltaVault par NetApp, Outscale fournit des services de Cloud Computing (IaaS) de Classe Entreprise, répondant aux exigences réglementaires et locales, aux entreprises souhaitant augmenter leur agilité business et déployer rapidement des business modèles à valeur ajoutée, en interne et pour leurs clients.

Avec 15% de son chiffre d'affaires consacré à la R&D, Outscale a choisi dès sa création de proposer une offre alliant exigence et excellence. Ce choix stratégique lui a permis de gagner la confiance des éditeurs de logiciels, startups et grandes entreprises en Europe, en Amérique du Nord et en Asie. Outscale a reçu la certification de sécurité ISO 27001-2013 sur l'ensemble de son infrastructure européenne.

## ADOPTÉZ LE CLOUD À LA SECONDE

En Septembre 2016, Outscale lance la facturation à la seconde sur l'ensemble de ses services Cloud. Cette nouvelle approche confirme sa volonté d'innover en continu afin de répondre au mieux aux besoins et attentes de ses clients, notamment en termes financiers.

En s'engageant comme le premier fournisseur de Cloud à mettre fin au paiement à l'heure, Outscale tient ses promesses d'offrir un Cloud plus intelligent, performant et facile à déployer, toujours au service de la croissance de ses clients.

## ELARGISSEZ VOS OPPORTUNITÉS BUSINESS ET DÉMARREZ VOS PROJETS BIG DATA

Afin d'accompagner les entreprises dans la réussite de leurs projets Big Data, Outscale propose une solution sur-mesure. Son offre Cloud On-Demand permet d'ajuster en temps réel les capacités de calcul et de RAM pour une infrastructure adaptée à ces nouveaux enjeux.

Outscale propose à ses clients une plateforme entièrement dédiée au Big Data et basée sur des technologies Hadoop et MapR. Cette suite d'outils leur permet de créer des applications hautement distribuées et fortement scalables, afin de traiter efficacement les gros volumes de données.

## VIVEZ LE BIG DATA SUR LE CLOUD OUTSCALE

- Grâce à une infrastructure Cloud reposant sur des solutions fiables et parmi les plus innovantes du marché : NetApp, Cisco, Intel et Nvidia.
- Avec un écosystème de partenaires technologiques pour vous accompagner sur tous vos projets Big Data.
- Avec notre eBook autour du Big Data que vous pouvez retrouver dès maintenant sur <https://pages.outscale.com/big-data>

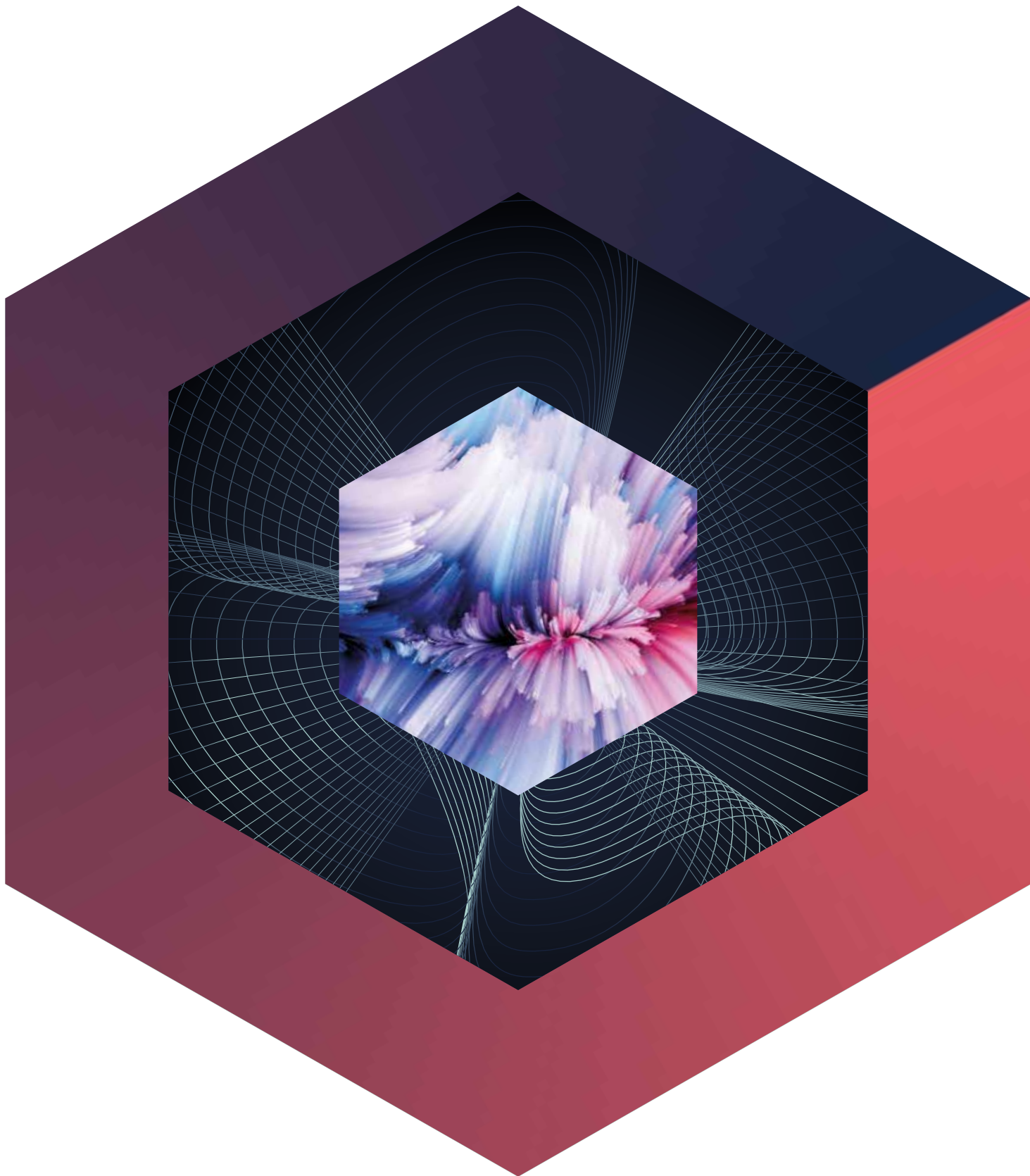


Créez, déployez et sécurisez  
votre application Big Data  
sur le Cloud Outscale.

Téléchargez notre ebook Big Data :  
[pages.outscale.com/big-data](http://pages.outscale.com/big-data)



Je suis un Cloud agile.  
Je réponds aux exigences liées  
à une nouvelle approche de vos  
données. Je vis le Big Data.  
Je suis Data Centric.



## PENTAHO

Urszula RADCZYNSKA-GAUTIER

Directrice Commerciale Grands Comptes

+33(0)6 73 14 34 14

ugautier@pentaho.com

Paris France



www.pentaho.com/fr



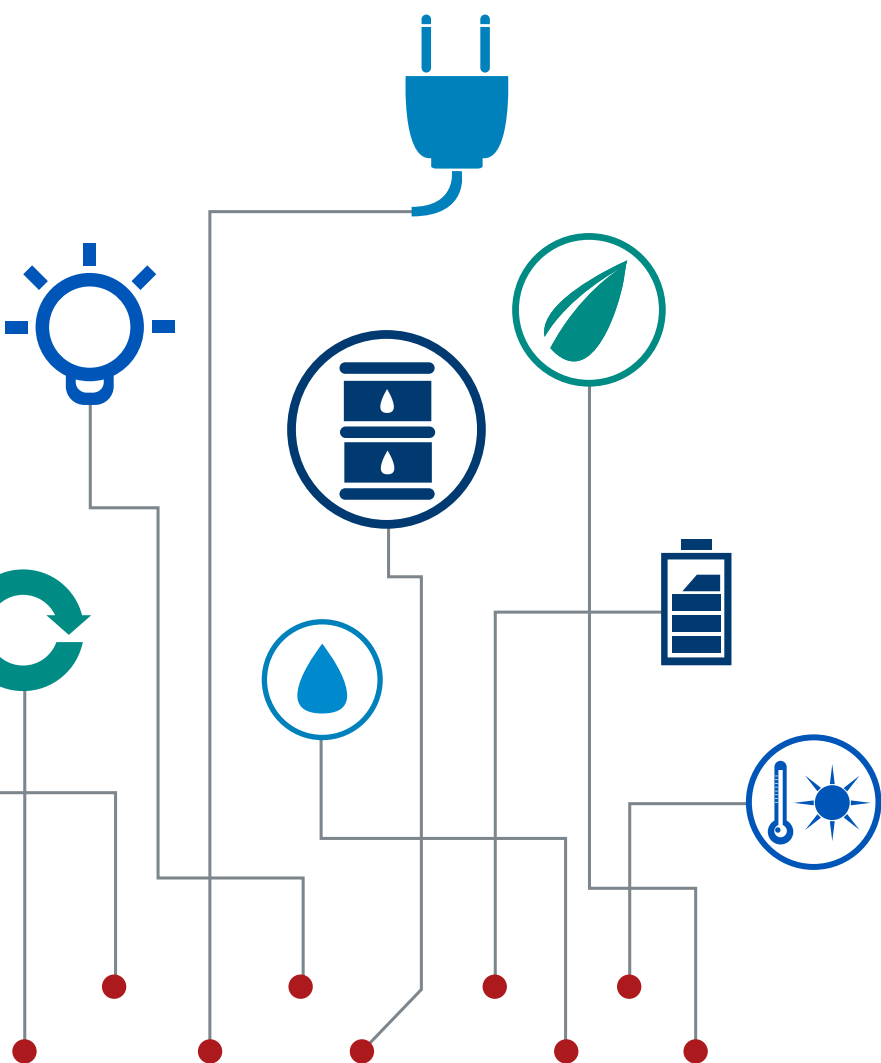
**PENTAHO, SOCIÉTÉ D'HITACHI DATA SYSTEMS, EST UN ACTEUR MAJEUR DE L'INTÉGRATION DE DONNÉES ET DE L'ANALYTIQUE AVEC UNE PLATEFORME OPEN SOURCE POUR L'ENTREPRISE QUI PERMET LE DÉPLOIEMENT DE DIVERS PROJETS BIG DATA.**

Sa plateforme unifiée et complète pour l'intégration et l'analyse des données est totalement intégrable et assure la gouvernance des données pour inclure des fonctions analytiques puissantes dans n'importe quel environnement. La mission de Pentaho est d'aider les entreprises de divers marchés à capter la valeur de toutes leurs données, y compris des Big data et des données issues de l'Internet des Objets, leur permettant ainsi d'identifier de nouvelles sources de revenus, d'opérer plus efficacement, de fournir des services remarquables et de minimiser les risques.

**Nouveauté :** Pentaho combine de façon native des fonctions d'intégration avec la visualisation et l'analyse de données. Grâce à la possibilité de visualiser les données en amont, les utilisateurs peuvent maintenant inspecter les données à la volée ainsi qu'à n'importe quel stade du processus de préparation des données. Parmi les fonctions clés de Pentaho Business Analytics :

- Croiser et virtualiser les jeux de données à la volée afin d'accélérer l'accès et accroître la flexibilité lors de l'agrégation des données
- Assurer en continu le croisement des données à grande échelle grâce à une « transmission optimisée », c'est-à-dire la transformation des données orientée vers les processus de traitement les plus efficaces
- Mettre facilement en forme les données les plus complexes
- Suivre et stocker les détails de traçabilité des données à chaque exécution de processus
- Améliorer les fonctions de collaboration de façon à partager les résultats d'exploration de données avec des outils perfectionnés de modélisation en ligne (modification du modèle en aller et retour)

**Pour évaluer gratuitement Pentaho Business Analytics :**  
<http://www.pentaho.fr/testdrive/>



Integrate.  
Blend.  
Analyse.

The Power of  
Big Data at Work

Learn more at [Pentaho.com](http://Pentaho.com)



### LES FABRICANTS DEVRAIENT AVOIR UNE APPROCHE PRAGMATIQUE DE L'INTERNET DES OBJETS (IOT)

Par Sebastien Cognet, Ingénieur commercial EMEA – Pentaho, a Hitachi Data Systems Company

Au siècle dernier, les mineurs de charbon emmenaient des canaris dans les tunnels de la mine pour détecter les niveaux de toxicité du monoxyde de carbone. Ce gaz dangereux aurait tué les canaris avant de tuer les mineurs, ce qui donnait ainsi l'alerte pour sortir immédiatement des tunnels. Aujourd'hui, les mineurs utilisent des capteurs de chaleur à combustion infrarouges et catalytiques pour détecter les gaz toxiques et inflammables. Reliés à Internet, ces capteurs permettent à un agent de sécurité de contrôler à distance et même de prévoir les risques potentiels avant qu'ils n'aient une chance de s'aggraver. Bienvenue dans une application révolutionnaire d'intervention et de maintenance prédictives !

Selon les prévisions d'IDC, le marché mondial de l'IoT atteindra 7,1 trillions de dollars d'ici 2020 avec des données qui représenteront 10% de toutes les données sur terre. Il est évident que de plus en plus d'entreprises veulent prétendre à leur part de cette manne de l'IoT, mais bon nombre d'entre elles ne savent pas vraiment par où commencer et comment définir leurs projets. Je lisais récemment que les producteurs de lait installent des capteurs sur les vaches pour détecter l'ovulation et les périodes où elles augmentent leur production de lait. Cependant, lire les données des capteurs n'est pas gage de réussite.

J'encourage les entreprises à adopter une approche pragmatique de l'IoT. Plutôt que de commencer par un vaste concept élaboré de l'IoT, je préconise de revenir à une problématique métier qu'il est véritablement nécessaire de résoudre. Par exemple, la ferme laitière pourrait examiner sa chaîne de production et découvrir que le principal problème n'est pas la production de lait, mais une altération et une perte. Dans ce cas, cela aurait plus de sens de déployer des capteurs pour préserver la fraîcheur du lait durant le transport et le stockage que de les utiliser pour augmenter la production.

**Une approche pragmatique, qui se concentre sur un seul problème métier concret à la fois, donne l'opportunité d'explorer et de résoudre certaines des difficultés techniques spécifiques à l'analyse des données IoT. En voici cinq des plus couramment rencontrées :**

- **Volume et "parasites" des données :** exploiter l'IoT implique souvent de trouver une aiguille dans une botte de foin. Cependant, selon le cas d'utilisation, les « objets » peuvent générer d'énormes volumes de données. Le défi consiste alors à filtrer les informations parasites et trouver celles qui sont vraiment importantes. C'est pourquoi de nombreuses entreprises y intègre une « analyse de flux » et une « analyse de processus ». La première fournit en temps réel des informations issues des flux de données telles que les parcours de navigation, logs, données de mesure et la seconde consiste à prendre des captures de données machine, plutôt que de libérer une avalanche de données en temps réel.
- **Urgence et latence :** l'IoT sous-entend différents niveaux d'exigence en termes d'urgence et de latence. Il est important d'en tenir compte parce qu'on s'attend à interagir avec le « monde réel » en temps réel, si bien que de nombreux événements appellent à une latence nulle. Ainsi, ce capteur dans la mine doit déclencher une alerte dès qu'il détecte la présence de gaz toxiques. Sinon, revenons au canari ! Egalement, d'autres informations de l'IoT ne sont peut-être pas nécessaires « juste à temps », telles que les données régulièrement recueillies pour continuer à affiner et améliorer le modèle prédictif lui-même. Ces données peuvent potentiellement être collectées et traitées plusieurs fois par jour, par exemple. Les architectures temps réel et de traitement par batch ont toutes les deux leur place dans l'IoT.
- **Variété de données provenant d'un manque de normes :** bien que les industries travaillent activement pour y remédier, un problème majeur de l'IoT est le manque de normes. Trop peu de normes gouvernent les diverses données générées par différents capteurs. Cela signifie que des appareils similaires de différents fabricants utilisent des formats de données complètement différents et génèrent des données à des fréquences différentes. Pour pallier l'immense variété de données existantes, une plateforme analytique doit être ouverte et indépendante d'un matériel.
- **Croisement de données :** dans le cas d'utilisation des opérations préventives par exemple, les entreprises veulent collecter les données provenant des « objets » (tels que les capteurs de chaleur à combustion infrarouges et catalytiques) et les croiser avec des données relationnelles pertinentes, telles que des accords de maintenance, informations de garanties et de cycle de vie des composants. Il est alors essentiel que les entreprises puissent se fier aux données à partir desquelles elles prennent leurs décisions importantes. Ce qui nous conduit au point suivant.
- **Traçabilité des données :** il s'agit essentiellement de l'enregistrement et de l'authentification de l'ascendance et de la véracité des données. Ceci est particulièrement important pour préserver la « santé » des données et fournir une chaîne de conservation des données vérifiable.

Une collaboration entre l'équipe informatique et des experts métiers sera plus que jamais essentielle pour l'analyse des données IoT. En plus de ceux qui comprennent les données, il faut des experts qui comprennent les données dans le contexte d'appareils ou de capteurs spécifiques. Alors que n'importe quel analyste peut comprendre les données dans le contexte d'indicateurs de performances d'une entreprise, seul un vétérinaire serait à même d'expliquer ce que des variations de température signifient en termes de fertilité et de production de lait.

Cependant mes conseils aux entreprises pour une adoption de l'IoT par étape, ne doivent limiter aucune ambition !



**ENJEUX**

Air Liquide est le leader mondial des gaz, technologies et services pour l'industrie et la santé. Présent dans 80 pays avec environ 68 000 collaborateurs, le Groupe sert plus de 3 millions de clients et de patients. Dans le cadre de sa démarche d'innovation Air Liquide a souhaité utiliser les outils de Proxem pour mieux identifier :

- les **découvertes scientifiques** à fort impact pour ses métiers actuels et ceux de demain ;
- les **nouveaux modes d'usages** de ses produits (oxygène, hélium...) ;
- les **leaders d'opinion** et **partenaires potentiels** au niveau mondial.

Dans cette perspective, Air Liquide a notamment sollicité Proxem en 2016 pour réaliser une cartographie du marché de la molécule d'oxygène.

**DÉFIS**

Dans l'ère des Big Data, la maîtrise de l'information est plus que jamais un enjeu stratégique pour prendre de meilleures décisions et piloter les entreprises. 80% des données produites par la révolution numérique sont de nature textuelle.

Les outils d'analyse de données chiffrées et les logiciels de veille classiques sont dépassés par cette abondance de **données textuelles**.

Les technologies d'analyse sémantique et d'intelligence artificielle deviennent indispensables pour **garder l'avantage** dans un marché concurrentiel mondial.

**PROJET**

Air Liquide recherchait un outil capable d'**explorer massivement le web public** afin d'identifier et d'organiser l'information publique disponible concernant les découvertes technologiques et les marchés potentiels de la molécule d'oxygène.

Proxem propose pour cela un logiciel d'analyse de données textuelles en mode SaaS. Utilisable de manière autonome, il place l'**expert métier au cœur** du processus d'analyse et de décision. Fondé sur le meilleur de l'intelligence artificielle et du traitement du langage, le logiciel Proxem **collecte et analyse** les documents issus du web en temps réel et dans toutes les langues. Il permet ainsi de repérer les thématiques pertinentes, quantifier les tendances et identifier les corrélations intéressantes par text mining et data mining.

**RÉSULTATS**

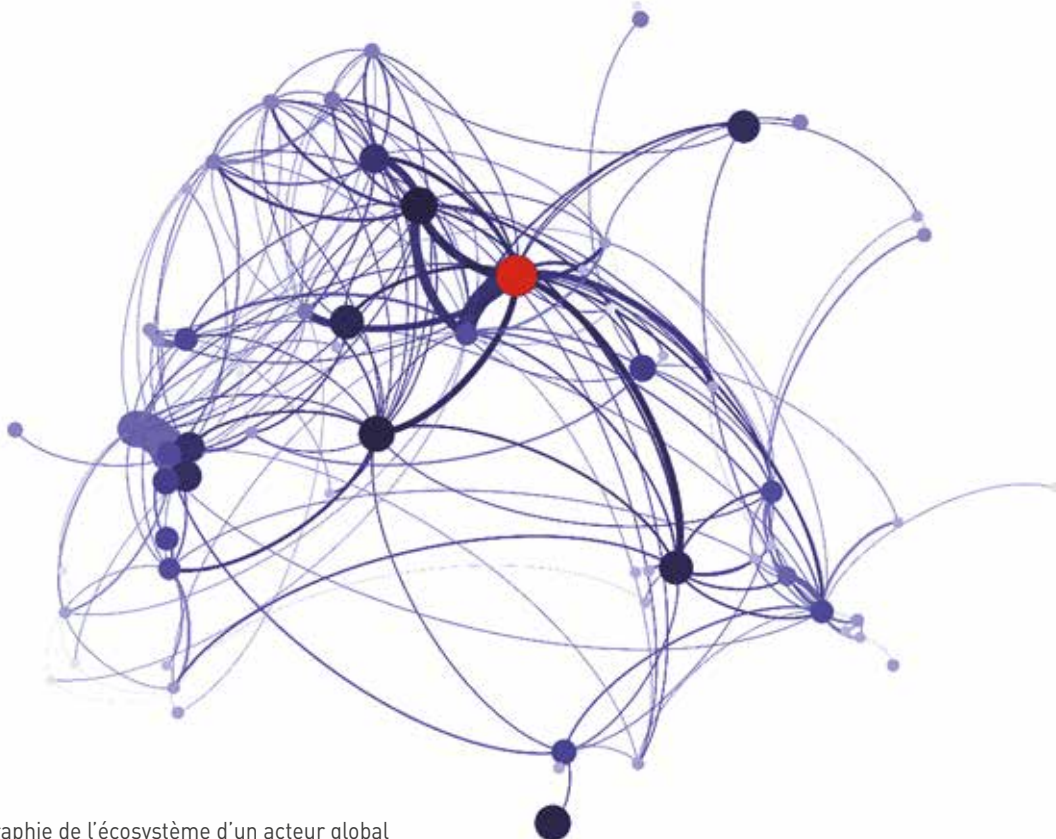
En **quelques semaines**, Air Liquide a pu identifier une cinquantaine de sujets d'intérêt pour le marché de la molécule d'oxygène, dont notamment :

- des **besoins croissants** en matière de traitement des maladies respiratoires chroniques ;
- des **technologies émergentes** d'utilisation de l'oxygène dans le domaine des transports ;
- des **méthodes** de conditionnement adressant un public de masse.

Sur le plan organisationnel, l'utilisation du logiciel Proxem a également permis à Air Liquide de mieux cerner des sujets de rupture. Dans une démarche **d'innovation**, elle complète ainsi efficacement des méthodes plus classiques comme le brainstorming et la veille traditionnelle.

**PERSPECTIVES**

Air Liquide multiplie aujourd'hui les projets d'analyse de données textuelles menés avec Proxem par différentes équipes. L'ambition de cette collaboration est de consolider les connaissances produites par les différents experts pour dégager une vision d'ensemble et faire de Proxem un des partenaires technologiques de la **transformation digitale** d'Air Liquide.



Exemple de cartographie de l'écosystème d'un acteur global

# Prendre de meilleures décisions avec l'analyse de données textuelles



- ✓ Mesure des forces et faiblesses
- ✓ Détection de risques et d'opportunités
- ✓ Cartographie de l'écosystème
- ✓ Application multilingue en temps réel



Content analytics



Market intelligence



Connaissance client



Ressources humaines



**DÉCOUVREZ TOUTE L'HISTOIRE QUI VIT DANS VOS DONNÉES**

A travers une approche unique, Qlik répond aux besoins croissants des utilisateurs et des entreprises : du reporting à l'analyse visuelle en libre-service en passant par l'analyse guidée, intégrée et personnalisée des données. La plate-forme d'analyse visuelle de Qlik aide les entreprises à optimiser leur prise de décision en combinant des données issues de différents systèmes. Grâce à une vue complète de l'activité, les entreprises peuvent mieux planifier, mieux gérer la complexité, et définir des plans d'actions efficaces. Qlik offre une excellente ergonomie et une grande facilité d'usage pour les utilisateurs, et répond aux nouveaux besoins de mobilité, sans compromettre la gouvernance ni la sécurité des données. En 2016, Qlik se classe parmi les leaders du Magic Quadrant de Gartner de Business Intelligence et des plates-formes d'analyse pour la sixième année consécutive, grâce à l'exhaustivité de sa vision et à sa capacité d'exécution. Qlik offre les capacités d'analyse les plus puissantes du marché.

**KING.COM SE TRANSFORME EN ANALYSANT LE COMPORTEMENT DES CLIENTS**

King.com, l'une des plus grandes entreprises européennes de jeux en ligne, avait besoin d'un moyen de définir des modèles dans la masse de données générées par les joueurs. Les jeux de King.com génèrent plus de 2 milliards de lignes de données par jour - et ces volumes s'accroissent chaque jour. Dans cette mer de données, se trouvent les secrets de la rejouabilité du jeu, la fidélisation des clients et la compréhension de l'efficacité des campagnes de marketing. Avant Qlik, les analystes commerciaux de King.com n'avaient aucun moyen facile d'extraire la valeur de toutes ces données, qui étaient stockées dans un système big data de type Hadoop. Les ingénieurs marketing devaient demander aux développeurs de jeux d'écrire des scripts dans le but de suivre le nombre de parties jouées par canal. « Un autre avantage majeur de l'utilisation de Qlik est la capacité à extraire les données de différentes sources », poursuit Mats-Olov Eriksson, architecte de données chez King.com.

**Analyse du comportement et du retour sur investissement du marketing**  
Grâce à Qlik, les analystes commerciaux chez King.com sont maintenant en mesure d'analyser le comportement de jeu de 40 millions de clients. Pour la première fois, ils sont capables de déterminer le retour sur investissement de leurs campagnes de marketing. Ils ont désormais une visibilité sur les mesures - comme le nombre de joueurs, le nombre de parties jouées, la durée de jeu, le revenu moyen par jour et par utilisateur, les rétentions continues du deuxième jour, les utilisateurs actifs par mois et par jour et plus encore.

**Visualiser les relations dans d'énormes volumes de données**  
King.com a choisi Qlik pour l'expérience associative, ce qui a permis aux utilisateurs métier d'explorer ce gros volume de données de jeux afin de trouver les informations dont ils avaient besoin. Avec Qlik, les utilisateurs ne sont pas obligés de suivre des protocoles prédéfinis. Ils peuvent naviguer et interagir avec les big data de la manière qu'ils souhaitent. Plus important encore, grâce à la fonction associative unique de Qlik, ils sont à même de réellement visualiser les liens existants dans le Big Data et de tirer profit de toutes les dimensions en réalisant différentes combinaisons au cours de l'analyse.

Plus de 100 applications Qlik sont maintenant utilisées quotidiennement chez King.com, à des fins très variées. « La mise en œuvre de Qlik a coûté 20% moins cher que les autres solutions. La rentabilisation s'est faite en seulement quelques mois. »

**38 000 clients satisfaits**

*La meilleure preuve de l'efficacité de Qlik ? Nos équipes marketing ne peuvent plus s'en passer ! C'est devenu un outil de pilotage indispensable pour nous.*

**Antoine Lacharmoise**  
Directeur CRM & BI, Photobox



*Les capacités de Qlik permettent à nos clients de piloter leur supply chain en toute autonomie. C'est un avantage concurrentiel fort pour arvato.*

**Philippe Le Coq**  
IT Manager, arvato



*La dataviz offerte par Qlik Sense est un élément essentiel du processus prédictif : elle permet de redonner le pouvoir à l'utilisateur. »*

**Claudio Borlo**  
DSI, Rexel



*Qlik repousse les limites de la BI traditionnelle grâce à la puissance de son moteur associatif et à la rapidité des développements.*

**Christophe Mébarek**  
Responsable SI du domaine Soutien et Services Client, Thales  
Systèmes Aéroportés





TOURISME

COÛTS DES ETUDES

Qlik® révèle des relations insoupçonnées entre vos données.

Qlik révèle les connexions entre les informations que d'autres solutions d'analyse n'ont pas la capacité à détecter. Notre modèle associatif unique, dévoile l'histoire qui se cache dans vos données, pour vous permettre de prendre de meilleures décisions.

© 2016 QlikTech International AB. All rights reserved.

**Qlik**   
[qlik.com/wholestory](http://qlik.com/wholestory)

**Jessica DOYEN**

Responsable Événementiel & Relation Client

+33(0)1 80 96 00 00

[jdoyen@report-one.com](mailto:jdoyen@report-one.com)

11 Villa Gaudalet 75011 Paris France



[www.report-one.fr/](http://www.report-one.fr/)

# Report One

Pilotez vos processus métiers !



## REPORT ONE ACCOMPAGNE SES CLIENTS SUR DES PROJETS DE PILOTAGE DE LA PERFORMANCE ET MANAGEMENT DES PROCESSUS MÉTIERS, AVEC DES SOLUTIONS DE BI ET CRM.

Avec notre offre de BI orientée business reporting, MyReport, aide les Directions opérationnelles à combler les « 100 derniers mètres » entre les données contenues dans leur système d'information corporate et leurs besoins en reportings. MyReport permet de rendre la main aux utilisateurs sur le flux d'informations et d'obtenir une vision agrégée, fiable et unique.

En tant qu'éditeur conseil, nous mettons en œuvre nos solutions de business reportings dans le cadre de projets complets : de l'audit des besoins au transfert de compétences avec l'accompagnement de notre réseau d'intégrateurs de proximité et d'experts métiers. C'est le cas pour Terre d'Alliances, société coopérative agricole de collecte et d'approvisionnement en grandes cultures, élevage et vigne. Leader en région Rhône-Alpes avec plus de 80 ans d'existence, le groupe compte 2 580 adhérents et atteint le chiffre d'affaires consolidé 2014/2015 de 225 millions d'euros. Terre d'Alliances est présente dans le secteur du transport de marchandises via sa filiale Ceretrans, ainsi que dans la logistique et le stockage de matières dangereuses via sa filiale Cérégrain Distribution.

Le Groupe souhaitait construire rapidement des tableaux d'analyses et de synthèse, tout en croisant les données de ses diverses applications professionnelles pour gagner en visibilité et réactivité dans le pilotage de ses activités. Les sources de données sont issues de systèmes et langages informatiques différents et installées sur plusieurs serveurs :

- Oracle pour la gestion de la production de semences et le back office des magasins Gamm Vert,
- AS400 pour la facturation, la gestion et la comptabilité,
- SQL pour le suivi des contrats de céréales et des situations de marchés de céréales,
- Sage pour les RH...

En l'absence d'entrepôt de données, les collaborateurs peinaient à réunir facilement et automatiquement toutes les informations nécessaires pour l'analyse et manquaient de souplesse dans la gestion et l'administration de leurs données.

Avec l'accompagnement de Business Cairn, partenaire de notre réseau d'intégrateurs de proximité, Terre d'Alliances met en place la solution de business reporting MyReport.

## OBJECTIFS :

- Créer un entrepôt de données permettant une meilleure gestion des données de l'ensemble des activités du Groupe,
- Croiser les données issues de différents logiciels (Paie, RH, ...) et supports dans une même analyse pour les besoins de la Direction jusqu'aux opérationnels,
- Diffuser aisément les rapports au sein de la société.

Après une formation rapide, les premiers tableaux sont créés. **Et un mois après l'intégration de MyReport, Terre d'Alliances devient autonome dans son utilisation de la solution** : suivi d'activités, quantités vendues, commandes clients par produits et par zones commerciales, etc. Les nombreux tableaux de bord sont désormais mis à jour et prêts à être exploités à chaque début de journée.

La grande plus-value de l'utilisation de MyReport : **puvoir consolider les données de différents univers, quelque soit la source, pour améliorer la précision des reportings et affiner les prévisions.** Finis les différents fichiers à regrouper. MyReport interroge les applications professionnelles et délivre rapidement le tableau de bord demandé. Terre d'Alliances peut désormais récupérer les données de toutes ses applications et les agréger pour obtenir des tableaux classiques (suivi d'activité, marge, commandes, prix moyens...), mais également des analyses plus poussées croisant les données de production, comptables, RH et boursières.

**Les utilisateurs peuvent maintenant gérer leurs données en toute autonomie et produire des reportings inédits, en toute simplicité.** Ils sont extrêmement satisfaits d'obtenir facilement des chiffres pertinents et fiables, d'être capables de réaliser de meilleures prévisions, de suivre l'activité de leur service au quotidien et ainsi agir pro activement sur la gestion des risques. Ils utilisent également le module d'envoi de mails automatisés pour communiquer des tableaux de bord journaliers ou hebdomadaires de manière ciblée avec gestion des alertes.

Simple, fiable et économique, la solution MyReport a transformé les processus et méthodes de travail de Terre d'Alliances, qui jouit d'une autonomie totale dans la collecte, la consolidation, et le croisement de ses données. La construction de tableaux d'analyses et de prévisions depuis Excel® a pris une nouvelle dimension. Les tableaux de bord dynamiques diffusés sur le web et statiques sont améliorés. Le partage de l'information entre les différents services est valorisé. Et surtout : l'ouverture du champ des possibilités en matière de reportings avec des recoupements inédits, doublée d'une **économie de temps appréciable, permettent à chaque collaborateur de se concentrer sur l'analyse et la prise de décision.**





### AUJOURD'HUI QUELLE EXPERTISE L'ÉDITEUR REPORT ONE PROPOSE-T'IL AUX ENTREPRISES CONCERNANT LE BIG DATA ?

Nous sommes un éditeur conseil de solutions logicielles. Notre vocation est d'apporter des solutions de pilotage de processus de management simples et accessibles pour un large public BtoB. Notre but étant de démocratiser le reporting au sein de l'entreprise, et de fournir à chacun les outils et bonnes pratiques pour piloter son activité simplement et en toute autonomie. Nous développons des technologies rapides à intégrer, économiquement abordables, à même de prendre en compte tous types de données : du simple fichier Excel® à la base de données, en passant par les logiciels métiers les plus sophistiqués. Les utilisateurs métiers peuvent ainsi reprendre la main sur leurs données – quelles qu'elles soient – et le pilotage de leur service.

### QUEL EST VOTRE CONSTAT SUR LE MARCHÉ DU BIG DATA ?

Il est très simple : aujourd'hui une multitude d'acteurs se positionnent sur ce marché. Et ce, aussi bien des acteurs historiques que des challengers, qui tentent tous d'apporter une solution à travers des outils encore trop difficiles à prendre en main par les utilisateurs métiers, souvent les premiers concernés par ce type de besoin. Chez Report One, nous nous sommes posés la question de l'exploitation du big data au sens le plus opérationnel possible : comment optimiser de façon efficace la gestion des données et des processus métiers ? Nous avons donc privilégié dans notre réflexion stratégique et réflexion produits, les acteurs métiers de l'entreprise afin de leur proposer les solutions les plus adaptées pour l'exploitation des différentes sources de données de l'entreprise.

En effet, notre constat est que les solutions mises en place au sein de l'entreprise – souvent via un investissement conséquent – se heurtent souvent sur le terrain au manque d'autonomie et de connaissances des équipes métiers. Ils n'ont pas une parfaite maîtrise des outils ainsi mis à leur disposition et destinés à leur donner accès au big data. À ce titre, nous répondons précisément à la question de savoir comment combler les « 100 derniers mètres » entre les solutions en place et le besoin concret des opérationnels au quotidien. Dans cet environnement de big data, un des éclairages qui nous semble donc majeur est de permettre aux acteurs métiers de piloter leur processus métiers aisément.

### QUELS SONT LES PRINCIPAUX BESOINS CONSTATÉS CHEZ VOS CLIENTS ? PARMIS QUELS SECTEURS D'ACTIVITÉ ?

L'attente principale est de faire face au déficit de pilotage des directions métiers. Nous sommes face à des entreprises qui s'organisent de manière centrale et standardisée pour mettre à jour les reportings des différentes activités. Il faut donc donner les moyens aux équipes de pouvoir approfondir les analyses en fonction des paramètres qui leur sont propres.

Par ailleurs, notre application est généraliste dans le sens où, elle s'adresse à tous types d'organisations et à toutes les fonctions : direction générale, direction financière, direction commerciale, direction marketing, direction industrielle, direction de la qualité, fonctions support, achat, logistique, etc. Quel que soit le domaine, nous nous appuyons sur des équipes expertes qui mènent des missions d'assistance et de maîtrise d'ouvrage visant à faire converger le côté standard de l'application avec les spécificités des métiers.

### QUELS SONT LES PRINCIPAUX AVANTAGES APPRÉCIÉS PAR VOS CLIENTS ?

Les clients apprécient les solutions de Report One pour leur autonomie de fonctionnement, leur capacité à s'adapter à tous les niveaux de compétences, leur robustesse, leur simplicité d'interface, et enfin pour leur capacité à répondre à de véritables enjeux économiques. Nous garantissons, et ce n'est pas si fréquent sur ce secteur, des projets dont le ROI est souvent inférieur à 18 mois.



**PRÉDIRE VOS VENTES ET OPTIMISER VOTRE SUPPLY CHAIN GRÂCE À NOTRE PLATEFORME BIG DATA**

**Une plateforme prête-à-l'emploi.**

Saagie accélère les projets Big Data en libérant ses clients des barrières technologiques et en leur permettant d'exploiter pleinement le potentiel de leurs données.

Comment ? Grâce à une plateforme Big Data prête-à-l'emploi, de bout en bout et orientée métier. De l'extraction à la data-visualisation en passant par le traitement ou le stockage, nous couvrons toute la chaîne de valorisation de la donnée en nous concentrant sur sa valeur métier ; pour qu'elle devienne un outil d'aide à la prise de décision stratégique en entreprise.

**Un exemple dans le monde des PME.**

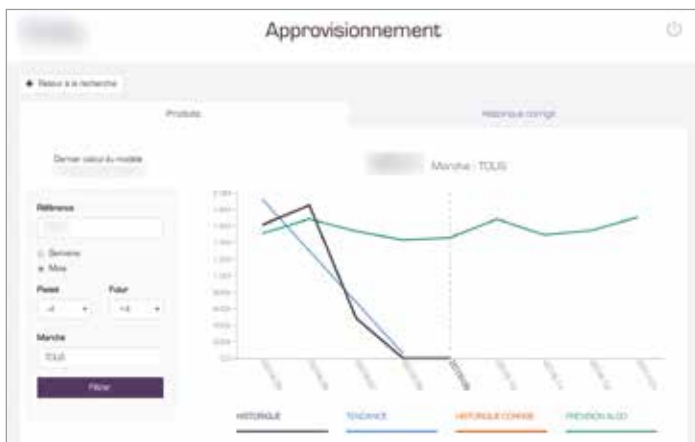
Didactic est l'une de nos références prouvant que le Big Data n'est pas l'exclusivité des grands groupes mais bel et bien l'affaire de tous. Preuve en est que les sujets traités ci-après (prévision des ventes et optimisation de la supply chain) et l'évolution de la demande client ont une portée quasi-universelle.

Didactic est un fabricant de dispositifs médicaux à usage unique (perfusions, gants, seringues...) pour des établissements de santé publics et privés français. Il intervient sur l'ensemble de la chaîne de valeur, de la fabrication en sous-traitance jusqu'à la commercialisation en passant par la stérilisation.

Confronté à des problématiques de gestion des stocks, ayant un délai d'approvisionnement à 5 mois et un délai de commande et livraison de ses clients à 2 jours, les moyens techniques existants ne lui permettent aujourd'hui que de générer des prévisions de ventes à une fréquence annuelle. Avec un chiffre d'affaires d'environ 40 M€, optimiser la supply chain est donc devenu un enjeu financier plus qu'important.

Dans ce cadre, Didactic a fait appel à nos services afin de mettre en place une plateforme Big Data permettant un suivi de la prévision des ventes plus fiable et ainsi d'optimiser les stocks et d'automatiser la construction du budget.

Pour ce faire, nous nous sommes engagés à lui apporter le meilleur des technologies Big Data afin de lui permettre de se détacher des problématiques de collecte et de traitement des données pour se concentrer sur l'apport de solutions concrètes aux problèmes rencontrés par ses équipes et ses clients.



Aperçu d'une smart App créée via Saagie Manager  
Interface de visualisation des prédictions pour les approvisionneurs

La priorité pour Didactic était de rendre accessible à ses commerciaux l'ensemble des informations relatives à un client et cela en itinérance. Les sources de données étant multiples (CRM, ERP, WMS, ventes, comptabilité, catalogue produits, open data...), elles ont toutes été intégrées au lac de données Saagie et ont permis la construction d'un tableau de bord sur-mesure permettant d'accéder à toute la granularité de l'information client de façon fluide (l'application est responsive et disponible en mode in et off-line) et ainsi d'optimiser la relation client et la recommandation de produits.

Nous avons par la suite travaillé ensemble sur la création d'un outil de gestion de ses approvisionnements en temps réel. Pour cela, nous avons créé une application prédictive pour la prévision des ventes en créant un algorithme sur-mesure à partir d'ERPs existants et de fichiers provenant d'anciennes versions d'ERPs afin d'arriver à une profondeur historique suffisante, et ainsi d'aboutir à un degré de probabilité bien au-delà de 95% pour les articles les plus vendus.

**Les utilisateurs aux manettes.**

Cette application intelligente offre également aux équipes de Didactic la possibilité d'ajuster au quotidien les volumes des différentes commandes ainsi que d'autres facteurs pouvant influencer sur les ventes ; ce qui permet ainsi à l'algorithme de se nourrir de ces ajustements et de mettre à jour en temps réel les prévisions associées.

A partir de ces prédictions de ventes, Didactic est désormais capable de visualiser l'ensemble des commandes par produit et ainsi d'optimiser sa supply chain tout en augmentant la satisfaction client grâce à une mise en adéquation de l'offre et de la demande.

Comme pour Didactic, Saagie offre à toutes les entreprises de gagner un temps non-négligeable au démarrage de leurs projets Big Data en leur permettant de bénéficier des derniers frameworks analytiques et en maximisant leur ROI. Mais c'est finalement l'agilité de la mise en œuvre, l'implication et la mixité des équipes (métier, SI, data) qui feront le succès de la démocratisation du Big Data en entreprise.



Frédéric Viguié, CEO de Didactic



# Saagie®

## Réinventez votre business avec Saagie

Nous accélérons vos projets Big Data grâce à une plateforme prête-à-l'emploi, de bout en bout et orientée métier.

### Ce que vous avez à y gagner ?

- ✓ Un accès simplifié aux technologies (Data Technology as a Service)
- ✓ Un potentiel data pleinement exploité
- ✓ Un retour sur investissement rapide



### Exemple de cas d'usages



Maintenance prédictive



Prévision des ventes



Segmentation du marché

### En savoir plus...



Par email  
[contact@saagie.com](mailto:contact@saagie.com)



Sur notre site  
[www.saagie.com](http://www.saagie.com)



Sur twitter  
[@saagie\\_io](https://twitter.com/saagie_io)

**Serge BOULET**

Directeur marketing

+33(0)1 60 62 11 11

comsas@fra.sas.com

Domaine de Grégy - Grégy-sur-Yerres  
77257 Brie Comte Robert cedex France



www.sas.com/france



## SNCF GARES & CONNEXIONS - LE BIG DATA ENTRE EN GARE

### SNCF GARES & CONNEXIONS ANALYSE LES FLUX DE VOYAGEURS DANS LES GARES AFIN D'AMÉLIORER LA QUALITÉ DU SERVICE RENDU AUX VISITEURS. LA GARE DU FUTUR SE REDÉFINIT AVEC LE BIG DATA ET AVEC SAS !

SNCF Gares & Connexions est la branche du groupe SNCF chargée de l'exploitation, de la maintenance et du développement des 3000 gares ferroviaires du réseau français. « 10 millions de visiteurs transitent par les gares chaque jour, » explique Marie-Caroline Bénézet, directrice digital. Dans ce flux colossal, des gens arrivent, partent, attendent, se restaurent, cherchent une correspondance... Et certains ne prennent pas le train : ils accompagnent un proche, utilisent des services en gare ou font des achats dans les galeries commerciales. La gestion des gares n'est plus seulement un métier ferroviaire !

#### La « connaissance client » entre en gare

Pendant des années, SNCF Gares & Connexions ne disposait pour piloter son activité que des données de trafic transmises par les transporteurs – auxquelles s'ajoutaient ponctuellement des comptages aux entrées des gares et des enquêtes menées auprès des usagers, par exemple en amont de travaux de rénovation ou de développement.

Depuis plusieurs années, la branche développe une position plus neutre par rapport aux transporteurs. « La connaissance client devient un enjeu de SNCF Gares & Connexions, » souligne Marie-Caroline Bénézet. « Avec le potentiel qu'offre le digital, nous travaillons à développer un lien avec le client le plus fin possible, pour apporter une réponse adaptée à chaque visiteur – qu'il soit voyageur ou non, occasionnel ou régulier, pour le loisir ou professionnel... »

Cette parfaite connaissance des flux, c'est tout l'enjeu de DATA & Connexions.

#### Un portail pour partager la connaissance client

Développé depuis 2015, le portail DATA & Connexions vise à fournir l'information la plus riche et la plus précise possible sur tout ce qui concerne les flux de visiteurs dans leur gare.

Outre les données « traditionnelles » (trafic, enquêtes), la nouveauté concerne l'exploitation des « données wifi » des visiteurs, ainsi que les informations provenant de multiples capteurs connectés : autant de sources qui fournissent des indicateurs agrégés (ex. nombre moyen de personnes, temps moyens passés en gare) et qui peuvent servir à l'ensemble des métiers : programmation des travaux, gestion des espaces commerciaux, information des usagers...

« L'important, c'est que les métiers puissent exploiter les données pour leur activité ! » insiste la directrice digital. Après une étude approfondie des besoins et une première phase de test, la solution peut maintenant être déployée en mode industriel, en commençant par deux piliers : l'analyse des flux de visiteurs, et l'affichage en gare.

#### Pilier 1 : analyse des flux de visiteurs

Dans les grandes gares, l'exploitation des données wifi donne aujourd'hui une connaissance précise des flux, avec des indicateurs pour les quantifier et les caractériser. Il est ainsi possible d'analyser la répartition du temps passé dans la gare, de mesurer la densité sur une zone donnée tout au long de la journée, et de suivre les flux en distinguant ceux qui arrivent, ceux qui partent, les voyageurs en transit ou les non-voyageurs. Cette connaissance nouvelle permet notamment d'analyser le comportement des visiteurs, en situation normale ou en situation « perturbée » : retard, travaux, ou encore afflux de voyageurs en raison d'un événement particulier –Euro2016- ou saisonnier –départ en vacances-.

A terme, elle doit naturellement contribuer à optimiser l'aménagement des gares, y compris la signalétique, mais aussi de comparer les données d'une gare à l'autre, ce qui était naguère impossible.

#### Pilier 2 : optimisation de l'affichage en gare

Le deuxième pilier concerne toutes les informations données aux voyageurs en gare, notamment via les grands panneaux d'information. « Les données permettront de voir où nous sommes performants et où nous le sommes moins, selon l'horaire, le type de train ou le type de gare », explique Marie-Caroline Bénézet.

Un exemple concret : l'affichage du quai d'un train. Il est habituellement donné vingt minutes avant le départ. Que se passe-t-il si l'information est donnée plus tôt ? Cela éviterait peut-être certains effets d'agglutination sous les panneaux... avec un risque accru de changement de quai – donc de nouveaux flux importants de voyageurs... Avec DATA & Connexions, différentes formules pourront être testées, avec une analyse en temps réel de leur impact !

#### Vers une logique de Test & learn

L'analytique rend en effet possible une nouvelle logique de type Test & learn, où l'on peut expérimenter de nouveaux aménagements dans une gare et observer rapidement les résultats, pour corriger, affiner, et éventuellement généraliser.

Cette logique a déjà été appliquée dans le cadre de l'installation de portiques de sécurité : différentes formules ont ainsi été testées, pour voir laquelle présentait le « meilleur » effet possible sur les flux. Elle permettra aussi d'évaluer divers dispositifs d'information en cas de travaux : l'examen des flux permettra de voir rapidement si les messages sont bien compris par les visiteurs, pour corriger rapidement si besoin, et essayer les bonnes pratiques dans d'autres gares et situations comparables.

#### SAS Inside : manipulation et partage des données

Le portail s'appuie sur une infrastructure big data sur Hadoop au sein de la « Big Data Fab » du Groupe SNCF. Le choix de la solution technique s'est porté sur SAS® Visual Analytics pour la visualisation des données, les tableaux de bord et les indicateurs, avec SAS® Enterprise Guide® en complément des outils traditionnels du Big Data (HIVE, R, Spark) pour le traitement préalable des données brutes. L'avantage ? La capacité de traiter de très gros volumes de données, mais aussi la possibilité d'agréger ces données dans des rapports qui peuvent être aisément diffusés aux cadres comme aux agents opérationnels en gare.

Au siège de SNCF Gares & Connexions, une petite équipe peut aussi explorer la base pour répondre à des questions précises provenant du terrain.

#### Perspectives : vers toujours plus de temps réel

L'outil étant encore jeune, le premier enjeu pour les équipes de DATA & Connexions consiste à poursuivre l'alimentation du portail en exploitant les multiples sources de données issues des gares.

Un autre chantier concerne l'ergonomie du portail, pour en démocratiser largement l'usage. Sur le plan technique, enfin, Marie-Caroline Bénézet insiste sur le pilotage de la qualité des données : « Nous voulons développer une architecture data qui aille encore plus loin dans le temps réel – et du temps réel de qualité », conclut-elle.

GARES & CONNEXIONS



# Andrew Pease

Principal Business Solution Manager

## INTERVIEW



### L'ANALYTIQUE EN TANT QUE FONCTION STRATÉGIQUE : L'ESSOR DU MÉTIER DE DATA SCIENTIST

Les entreprises n'ont jamais eu accès à autant de données et la plupart admettent que l'analytique peut avoir des effets positifs. À cet égard, Amazon est toujours cité en exemple. Pionnière de l'analyse des données, la société a en effet lancé la fonctionnalité « Les clients ayant acheté cet article ont également acheté... ». Aujourd'hui, les recommandations d'Amazon sont basées sur la liste d'envies de l'utilisateur, les articles vus et les articles achetés par d'autres clients ayant un comportement d'achat similaire. L'analyse prédictive gagne ainsi en précision.

La récente étude mondiale de McKinsey, « The need to lead in data and analytics », souligne également la nécessité d'ouvrir la voie dans le domaine des données et de l'analytique. La majorité des personnes interrogées pensent que, dans les années à venir, leurs activités analytiques auront un impact positif sur le chiffre d'affaires, ses marges et son efficacité opérationnelle. Il n'est pourtant pas si facile de tirer des enseignements utiles de gigantesques quantités de données, puis de prendre des mesures précises. De plus en plus d'entreprises placent tous leurs espoirs dans les data scientists.

### ENCORE UN CONCEPT À LA MODE ?

Aujourd'hui la Silicon Valley regorge de data scientists et les entreprises se bousculent pour recruter les professionnels avec le bon profil, ce qui témoigne de l'engouement actuel pour l'analytique. Alors, qu'est-ce qui fait que les data scientists sont si recherchés ? « À mon avis, la science des données va plus loin que la simple analyse », répond Andrew Pease, Principal Business Solutions Manager chez SAS.

« Un data scientist utilise des techniques plus avancées pour identifier les enjeux commerciaux, recueillir les données pertinentes et publier des informations exploitables. Il est capable de découvrir des tendances dans les données et de formuler des prévisions significatives. »

Outre des connaissances en statistiques, le data scientist doit posséder des compétences diverses. « Les data scientists doivent avoir une forte envie d'apprendre, d'innover et d'améliorer les choses. Bien sûr, ils doivent trouver les données, les analyser, les interpréter et partager les résultats, mais s'ils ne posent pas les bonnes questions dès le départ, toutes ces données et le meilleur bagage statistique au monde ne leur seront d'aucun secours », fait observer Andrew Pease.

### RÔLE DES DÉCIDEURS

Bien que l'analytique arrive en tête des priorités des dirigeants, nombre d'entre eux ne font pas suffisamment bien passer le message dans l'entreprise. Selon l'étude de McKinsey, 38 % des PDG affirment piloter le programme analytique de leur entreprise, ce qui n'est confirmé que par seulement 9 % des autres cadres supérieurs. Ces personnes citent plus volontiers les DSI, directeurs marketing ou chefs de divisions opérationnelles comme responsables des initiatives en matière d'analytique.

« Peu importe le décideur aux commandes, du moment que l'analytique constitue une fonction métier stratégique. Jusqu'à présent, l'analytique a néanmoins été souvent perçue comme une fonction secondaire des services informatiques. Promouvoir l'analytique au rang de fonction stratégique est donc la première des bonnes pratiques à mettre en œuvre. Toutefois il faudra du temps pour définir des profils analytiques, et les personnes concernées devront sans doute commencer par prouver qu'elles peuvent réussir avant d'accéder à un poste de plus haut niveau. »

L'analyse des données étant stratégique, il est primordial d'internaliser les processus analytiques à un moment ou un autre », affirme Andrew Pease.

### ANALYTIQUE EXPLOITABLE

La plupart des grandes entreprises recrutent déjà des data scientists. À l'ère des big data et de la convergence des secteurs d'activité, les entreprises réalisent que les informations contenues dans une transaction sont encore plus précieuses que la transaction proprement dite. « À ce jour, le secteur financier compte déjà un grand nombre de data scientists. La grande distribution suit et, cette année, la demande de data scientists devrait également exploser dans le secteur industriel. Avec la révolution de l'Internet des objets (IoT), l'analyse de grandes quantités de données relevées par les capteurs va revêtir une importance capitale ». Le succès de la science des données en entreprise ne repose pas seulement sur des algorithmes, conclut Andrew Pease.

Selon l'étude de McKinsey, l'un des principaux obstacles à un programme analytique efficace est le manque de communication. Et Andrew Pease confirme : « Les data scientists doivent être à la fois capables d'interpréter les données et de communiquer leurs découvertes aux décideurs de leur entreprise. Ils doivent également rendre les mécanismes analytiques assimilables par les personnes concernées dans l'entreprise. Les techniques de visualisation sont par exemple d'une aide précieuse, sachant qu'une image vaut mieux qu'un long discours - ou que les lignes d'une feuille Excel, en l'occurrence. »

# SCALED RISK

**Edouard ESTOUR**  
Chief Marketing Officer

+32 471 467 496  
edouard.estour@scaledrisk.com  
71 Boulevard de Sébastopol 75002 Paris France

CONTACT



www.scaledrisk.com



## SCALED RISK

### SCALED RISK ACCELERE LA MISE EN PLACE D'UNE PLATEFORME DATA ANALYTICS AU SERVICE DE LA CREATION DE NOUVEAUX SERVICES AUX INVESTISSEURS POUR UN GRAND GROUPE DEPOSITAIRE.

Les services financiers et les services aux investisseurs connaissent une profonde mutation, où la concurrence entre groupes bancaires s'accélère au contact des nouveaux acteurs de la fintech, de l'apparition des nouvelles technologies Big Data et Blockchain. Aussi, l'inflation réglementaire redessine la chaîne de valeur, et pose une forte pression sur les métiers et les systèmes d'information.

C'est dans ce contexte que le client de Scaled Risk, un grand dépositaire français, a souhaité mettre en place des services digitaux innovants au cœur de son offre de services aux fonds d'investissements. L'enjeu, outre de réussir cette transformation digitale, est de fournir aux métiers les moyens de créer et de faire évoluer ces nouveaux services, avec toute l'agilité, la réactivité et l'évolutivité requises.

Après la réalisation de POCs afin de tester la capacité des technologies Big Data à répondre à ces challenges, le dépositaire a cherché une solution Big Data unique capable de fournir une plateforme centrale, extensible et ouverte intégrant des fonctions de Data Management puissantes et flexibles pour la modélisation métier des données, et des fonctions de Data Analytics. Et après un processus d'appel d'offres, Scaled Risk a été choisi pour son caractère packagé et unifié, son architecture ouverte et évolutive, et ses fonctionnalités d'audit et de traçabilité intégrées.

La plateforme Scaled Risk permet l'agrégation de données multi-métier, d'appuyer des calculs variés sur ces données, et de mettre à disposition des gérants et des clients, les résultats de l'ensemble via des outils de construction de rapports et des outils de visualisation des données. Les gérants peuvent agréger, transcrire, appareiller en données référentiels (rating, market data, etc.) tout ou partie de l'historique de leurs inventaires (plusieurs milliards de lignes) et consulter le résultat directement dans une interface graphique mettant en avant des représentations visuelles variées. Enfin, la plateforme permet l'étude du comportement des investisseurs en fournissant aux gérants les moyens de chercher des corrélations entre les ordres de souscriptions/rachats sur les fonds au passif et des caractéristiques des référentiels.

Scaled Risk a déployé sa plateforme de Data Management et Data Analytics et a su répondre aux contraintes fortes de délai de mise en œuvre et de lancement des services : le lac de données a été mis en place en seulement quelques semaines, centralisant plus de 50 sources de données issues du système d'information existant, sans impacter ce dernier, sans casser les modèles et les processus métiers existants.

Ensuite, la modélisation des données métier de manière extrêmement flexible au sein de l'application, en réconciliant diverses sources hétérogènes, a permis d'apporter une réponse rapide pour la création des services, d'adapter ceux-ci très rapidement aux demandes du métier et de la réglementation, et de les faire évoluer à moindre coût.

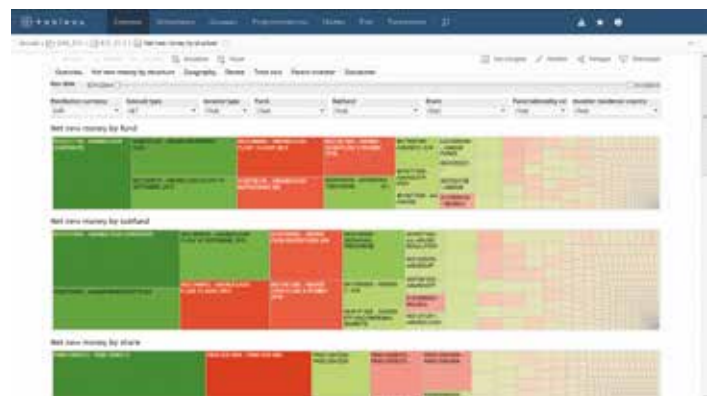
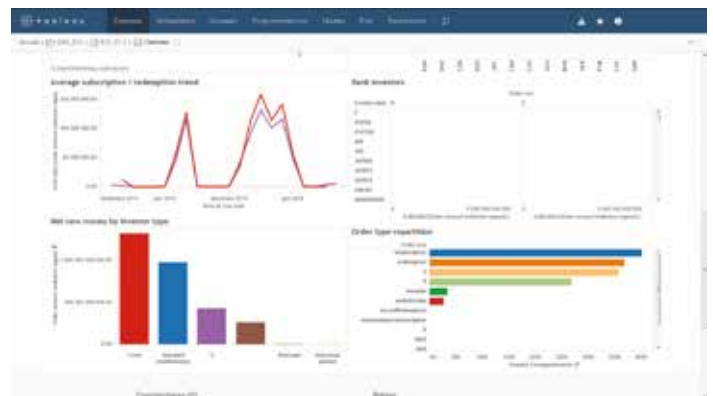
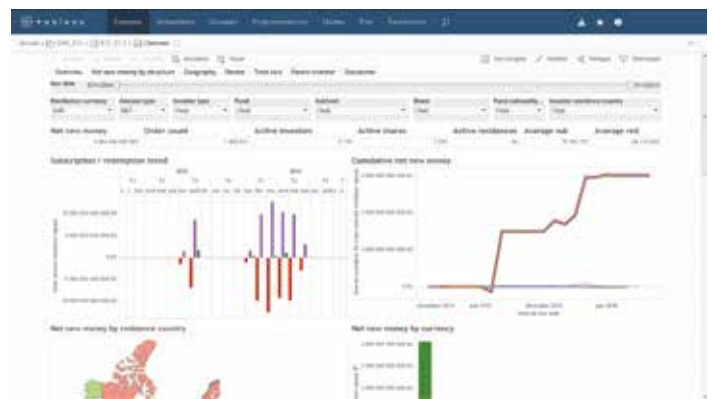
La solution intégrée et industrielle Scaled Risk assure également flexibilité, sécurité, audibilité, traçabilité, continuité de service, et une grande connectivité et ouverture aux applications standards. En outre, la gestion de la qualité des données est au cœur de la solution Scaled Risk.

Les équipes du client ont pu faire lever sur les capacités du Big Data en terme d'analyse et de flexibilité pour créer des services innovants d'étude du comportement des investisseurs, ou pour améliorer la production des rapports réglementaires. Le dépositaire a pu également mettre à disposition de ses clients investisseurs, directement dans son portail existant, des services de visualisation des données extrêmement puissants

par l'intégration native de Tableau Software avec la plateforme de Data Management Scaled Risk.

Scaled Risk commercialise ainsi des solutions innovantes de Data Management, de Data Analytics et de Risk Management dédiées aux métiers de la gestion d'actifs, des services investisseurs et services titres.

Plus d'informations sur [www.scaledrisk.com](http://www.scaledrisk.com)





« D'un côté, on demande au data management d'être assez flexible pour englober la quantité des nouvelles demandes business et réglementaires, mais de l'autre, on exige de lui toujours plus de structure et de gouvernance. C'est la quadrature du cercle ! »

### QUELS SONT LES CHALLENGES QUE RENCONTRENT LES FONCTIONS RISQUE ET CONFORMITÉ DANS LA GESTION DE LEURS DONNÉES ?

Scaled Risk est né du constat qu'aucune solution actuelle de data management ne pouvait répondre au défi posé par la vague de réglementations introduites après la crise de 2008. Calcul de risque de marché ou de liquidité, connaissance client, lutte contre le blanchiment, le financement du terrorisme et la corruption ... En moins d'une dizaine d'années, institutions financières et industriels se sont retrouvés confrontés à une explosion de nouveaux reportings, tous plus complexes les uns que les autres. Or la nécessité de collecter, d'agréger et d'analyser autant de données s'est très vite confrontée aux limites de SI vieillissants et déjà surchargés.

### LES ENTREPÔTS DE DONNÉES BASÉS SUR HADOOP PEUVENT-ILS APPORTER UNE RÉPONSE ?

En effet, au même moment l'on a vu apparaître les premiers projets de « data lake » basés sur les technologies big data type Hadoop. Souvent portés par des équipes marketing / produit, ces projets ont rapidement intéressé les fonctions risque : bien moins coûteux et plus scalable qu'un datawarehouse traditionnel, le data lake Hadoop semblait la solution idéale pour centraliser toutes les données nécessaires aux reportings. Malheureusement, pour la plupart, ces projets ont échoué. Pourquoi ? Tout simplement parce que la capacité d'Hadoop à gérer des volumes importants de façon flexible repose sur des logiques de dénormalisation (stockage « à plat », puis re-structuration à la lecture). Or si cela ne pose aucun problème dans un contexte d'analyse et d'expérimentation, il en va différemment dans un contexte réglementaire : dictionnaire de données, piste d'audit, « data lineage », gestion des versions, historisation : tous ces éléments, indispensables pour satisfaire les exigences de n'importe quel régulateur, sont absents d'Hadoop.

### EN QUOI L'APPROCHE DE SCALED RISK EST-ELLE DIFFÉRENTE ?

Depuis sa création en 2012 par des vétérans du risk management, la mission de Scaled Risk a été d'offrir aux institutions financières et aux industriels les extraordinaires performances des technologies big data tout en garantissant les qualités de gouvernance et de robustesse des systèmes traditionnels. Avec une obsession : augmenter le time-to-market des projets d'analyse et de reporting. Aujourd'hui, Scaled Risk est l'un des seuls acteurs à proposer une data management platform industrielle permettant de profiter du meilleur des deux mondes. Au centre de la plateforme se trouve une technologie hybride de modélisation des données prenant appui sur la base NoSQL HBase. Capable de répliquer et d'intégrer les multiples modèles des bases source, la DMP Scaled Risk permet également de créer à la volée des modèles métiers totalement logiques sans que ne soient affectés les modèles originaux. Cette approche, qualifiée « d'ELT », est particulièrement efficace et permet au data management d'acquiescer de l'agilité sans perdre en gouvernance. Par ailleurs, la traçabilité des données et des opérations est un aspect important de la solution. Rien n'est jamais écrasé dans Scaled Risk : la suppression d'une donnée ou sa mise à jour entraîne en réalité la création d'une nouvelle version, et la version précédente est toujours conservée. De même, la modification d'un modèle de données entraîne la création d'une nouvelle version de ce modèle. Ainsi le client bénéficie d'une piste d'audit complète et peut toujours avoir une vue du système dans l'état où il était à n'importe quel moment dans le passé (« as-of-date »).

### POUVEZ-VOUS NOUS DONNER UN EXEMPLE QUI POURRAIT ILLUSTRER CES CAPACITÉS ?

Oui, tout à fait. Scaled Risk a été retenu pour aider un grand groupe industriel à outiller un programme de conformité anti-corruption dans le cadre de la future loi Sapin II. Basé sur la DMP Scaled Risk, le logiciel a pu rapidement intégrer des données internes structurées (ERP, CRM, RH), non structurées (documents), ainsi que sur des données externes (Reuters), et mettre en place une batterie de 250 règles métier pour lever des alertes en temps réel sur des comportements susceptibles de présenter un risque de corruption. La direction de la conformité de ce client bénéficie de capacités d'audit extrêmement fines sur un historique illimité, toujours « live » et donc disponible à tout moment.

L'ensemble du développement, de l'intégration et du déploiement du logiciel sur un périmètre de plus de 10,000 employés en France et à l'étranger a été effectué en 3 mois par les équipes Scaled Risk et ses partenaires du groupe Neurones.

**Xavière TELLENT**  
CEO

+33 (0)6 73 804 708  
lively@scinetik.com  
78 - rue Taitbout 75009 PARIS France



www.scinetik.com

**SCINETIK**  
INTELLIGENCE + E-BUSINESS



Scinetik est un cabinet de conseil dont la mission est d'assurer pour ses clients, le développement pérenne et rentable de leur activité de commerce connecté.

## NOTRE DIFFÉRENCE : ASSOCIER DATA ET MARKETING

Notre mission est d'accompagner nos clients dans la définition et la mise en œuvre de stratégies de développement de leur business, sur l'ensemble des drivers : de la conquête à la fidélisation en passant par la conversion. Pour y parvenir et afin de s'assurer de l'efficacité des actions, Scinetik fonde l'ensemble de ses recommandations sur l'analyse de la donnée. Cette donnée vient de multiples sources : le webanalytic, les comportements d'achat, les données CRM, les données non structurées (notamment websemantique), les données publiques...

Nous croyons fondamentalement aux vertus de la data mais ne concevons pas notre métier sans y ajouter une dimension forte de stratégie et de marketing.

## DÉTAIL DE NOS PRESTATIONS

### Stratégie Business

- Mettre les nouvelles technologies, les nouvelles formes de communication et l'intelligence des data au service du développement du business des entreprises.

### Optimiser la performance du commerce connecté :

- Exploiter la donnée pertinente pour concevoir et optimiser les parcours cross-device & cross-canaux : audit de performance des interfaces & parcours; UX/UI design ; plan d'optimisation continue (A/B tests) ; audit de la performance des stratégies pluri-media pour optimiser les budgets investis.  
- Intégrer la digitalisation des magasins dans la réalité du business de chaque client

### Capter une donnée de qualité et assurer sa bonne exploitation :

- Mettre en place des systèmes de collecte de données (ex : plan de tag) et de stockage (DMP) afin d'en extraire l'intelligence (KPIs, algorithmes) en vue d'améliorer le pilotage des activités par les équipes métiers
- Concevoir et éditer de rapports multi-sources personnalisés et automatisés.

### Réinventer les conversations et le CRM

- Prendre soin de chaque client, prospect, ami pour développer la préférence durable à la marque et un bouche-à-oreille positif.
- Concevoir et mettre en place des stratégies autour du commerce conversationnel.

### Faire du Big Data un facteur de transformation de l'entreprise

- Appréhender le Big Data comme un levier de transformation durable de l'entreprise, ses métiers, son offre, ses prévisions de vente, sa relation à ses cibles et non comme un sujet technologique.

## NOS VALEURS : TRANSPARENCE – PRAGMATISME – CURIOSITÉ

- Transparence = accompagner nos clients dans leur montée en compétence.
- Pragmatisme = évaluer nos recommandations sur un ratio complexité-coût / apport de valeur et ne menons pas des projets simplement intellectuellement satisfaisants
- Curiosité = exploiter au mieux les technologies innovantes existant à travers le monde grâce notamment au foisonnement de startup

## SECTEURS CIBLES

Chez Scinetik, nous travaillons pour tous types de secteurs d'activités. Toutefois, notre histoire fait que nous travaillons majoritairement pour de grands groupes internationaux dans les domaines du retail, de la grande consommation, de la mode, de la décoration et du sport.

## FAIRE DU BIG DATA UN FACTEUR DE TRANSFORMATION DE VOTRE ENTREPRISE



Cadrage  
stratégique



Sources &  
dataquality



Modélisation  
statistique



Application  
métiers & partage

# LES BEST SELLERS SCINETIK



## 1 Comment prédire le succès d'un lancement de produit ?

Identifier les principales variables explicatives de la réussite ou de l'échec d'un lancement de produit pour détecter les signaux pertinents. Définir des seuils qui permettront de déclencher les actions correctives au plus tôt.



## 2 À quel moment déclencher des actions commerciales et avec quelle démarque ?

Croiser les données internes (stock, web analytics, CRM, ticket de caisse, ...) avec les facteurs explicatifs externes (météo, indice du moral des ménages, calendaires, géolocalisation, concurrence, etc.) pour décider de la période optimale pour lancer les opérations de promotion.



## 3 Comment l'écoute du web permet-elle de comprendre la réalité de la perception des marques ?

Utiliser les conversations du web social (forum, blogs, sites spécialisés, réseaux sociaux) pour mesurer les co-occurrences entre la marque et ses concurrents et évaluer les items d'image positifs et négatifs et leurs évolutions dans le temps.



## 4 En quoi les nouvelles sources de données permettent-elles de concevoir des services innovants ?

Enrichir la donnée client existante avec des données type circulation des individus dans les espaces de vente (et autres lieux de passage) pour créer des profils avancés. Élaborer de nouveaux services et offres personnalisés à partir de ce profiling.



## 5 Quelle modélisation du mix-media pour optimiser la rentabilité des investissements futurs ?

Auditer les campagnes pluri-media afin de construire un modèle d'attribution permettant d'améliorer l'efficacité des stratégies de communication.



## 6 Comment utiliser l'Open data pour mieux réinventer le magasin dans son contexte local ?

Réaliser une cartographie des zones étudiées en visualisant les performances des magasins et les facteurs explicatifs liés aux caractéristiques locales. Intégrer les outputs de l'analyse pour reconcevoir des magasins plus adaptés à leur environnement.



## SEENK LAB

CREATIVE INTELLIGENCE

Agence spécialisée dans le branding et le digital, Seenk a été amenée à travailler sur de nombreux sites e-commerce depuis sa création en 2000, pour des clients tels que Castorama, Darty ou Roland-Garros.

Dans ce domaine, la data nous apporte la capacité à mesurer l'efficacité de façon immédiate.

Seenk Lab, notre pôle innovation, représente l'extension de cette expertise en aidant les marques à connaître leurs clients et à repenser leurs canaux de ventes.

### SUIVRE LES CLIENTS ET PROSPECTS PARTOUT OÙ ILS S'EXPRIMENT

Nous vous aidons en particulier à :

**Connaître vos clients, partout où ils sont et s'expriment**, sur mobile ou sur ordinateur, sur votre site ou sur les réseaux sociaux...

**Définir des segments prédictifs de clients** : néophytes ou experts, avides de sécurité, de service clients ou attirés par le prix le plus bas, nous cernons vos clients pour définir des offres sur-mesure.

**Délivrer des recommandations efficaces**, dynamiques et qui s'adaptent au profil de vos clients, à leur historique et à leurs envies, pour vendre plus et accroître l'engagement client.

### TRANSFORMER L'EXPÉRIENCE D'ACHAT

Mais les marques sont aussi aujourd'hui désireuses de trouver de nouvelles façons d'interagir avec leurs clients et de retrouver, sur le digital, une relation client enrichie, favorisant l'attachement à la marque.

Seenk Lab associe sa connaissance de l'expérience client à celle de la data pour bâtir des interfaces et des parcours sur-mesure.

Nous optimisons les parcours d'information et d'achat, identifions les centres d'intérêt et profils des clients et concevons les sites, applications et interfaces sur la base de ces données.

### CRÉER DES ASSISTANTS PERSONNELS UTILISANT L'INTELLIGENCE ARTIFICIELLE

Nous nous fondons en particulier sur l'intelligence artificielle pour créer des assistants personnels et conseillers automatisés, capables de puiser leur savoir dans des sources externes (blogs, Wikipedia, presse...) et internes (catalogue, historique d'achat...) pour offrir à vos clients la meilleure réponse possible à leurs questions et optimiser les ventes.

Nous travaillons ainsi avec plusieurs acteurs qui souhaitent repenser profondément leur expérience client et leur capacité à prodiguer un conseil en ligne, dans des secteurs aussi divers que la mode, le vin, l'immobilier ou le transport.

### NOUS CONNAÎTRE

Seenk Lab est présent partout dans le monde avec des antennes à Paris, Boston, Francfort et Mumbai.

85 personnes constituent l'éco-système de Seenk Lab : data analysts, développeurs, designers et directeurs de créations, planneurs stratégiques...

### TECHNOLOGIE

Nous travaillons sur des applications utilisant :

- Des techniques sémantiques adaptées à chaque secteur pour comprendre le langage naturel de vos clients, leurs attentes et leurs besoins.
- IBM Watson, une technologie d'intelligence artificielle révolutionnaire apprenante, fonctionnant en langage naturelle et capable d'indexer des milliers de sources internes et externes. Celle-ci va transformer la navigation sur les sites e-commerce en comprenant directement le contenu que souhaite trouver l'internaute. Elle permet aussi de trouver la réponse la plus pertinente en analysant les tweets, blogs, posts Instagram ou sites extérieurs et en les reliant au catalogue de l'entreprise.
- Un algorithme prédictif propriétaire, nourri des données les plus opportunes pour prévoir aussi bien l'implantation de votre prochain magasin que le comportement de vos clients sur votre site.
- Une mesure de l'influence en temps réel. Capable d'identifier les influenceurs pertinents sur tous les réseaux et d'analyser leur ton, sujets de prédilection et traits de personnalité.

## FINDABILITY SCIENCES

### NOUS PARTONS DES DONNÉES



### NOTRE SOCLE TECHNOLOGIQUE

- UN ALGORITHME PRÉDICTIF PROPRIÉTAIRE
- UNE MESURE DE L'INFLUENCE EN TEMPS RÉEL
- IBM WATSON, UNE TECHNOLOGIE QUI VA CHANGER LE WEB

### POURQUOI NOUS SOMMES LÀ ?

- VOUS PERMETTRE DE CONNAÎTRE VOS CLIENTS, PARTOUT OÙ ILS SONT
- FAIRE DES RECOMMANDATIONS QUI MARCHENT
- CONSTRUIRE DES INTERFACES QUI CHANGENT LA VIE DIGITALE

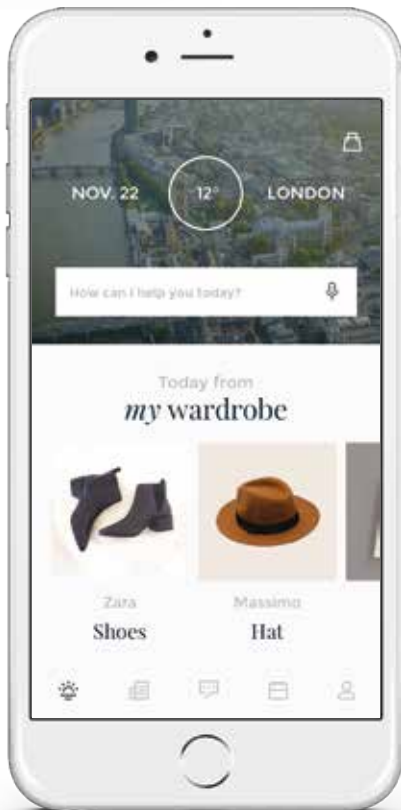


SEENK LAB

# INTELLIGENCE ARTIFICIELLE



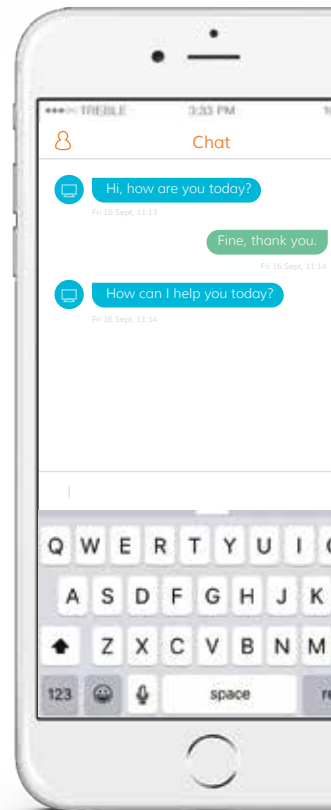
# PERSONAL SHOPPER



## WEARING



## WINE TASTING



## ...& more



**Isabelle REGNIER**  
Associé Fondateur

+33 (0)1 75 61 34 00 / +33 (0)6 76 09 82 46  
i.regnier@sentelis.com  
31 RUE DU PONT DE NEUILLY 92200 NEUILLY SUR SEINE  
France



www.sentelis.com



CONTACT

**SENTELIS INSPIRE LA TRANSFORMATION DU SYSTÈME D'INFORMATION DE SES CLIENTS, LES ECLAIRE SUR LES FONDATIONS DIGITALES POUR LA RÉUSSIR ET LES ACTIVE SOUS FORME DE CAPACITÉS INDUSTRIELLES PÉRENNES, QU'ELLES SOIENT TECHNOLOGIQUES ET/OU MÉTHODOLOGIQUES.**

Notre offre Big Data permet aux directions des systèmes d'information, qu'elles soient au début de leur parcours Big Data ou résolument engagées dans celui-ci, de valider dans leur contexte la valeur métier des technologies Big Data et d'en réussir l'intégration au cœur de leur système d'information.

Notre vision du Big Data s'inscrit en effet dans un changement de paradigme métier et système d'information où l'entreprise ne considère plus la donnée comme une simple denrée mais comme un capital stratégique. Cette évolution impose un renforcement de la gouvernance de la donnée et doit se transcrire dans l'architecture du système d'information de l'entreprise, qui lui-même doit devenir digital.

Notre modèle d'architecture 3.0 explicite ce modèle d'architecture digitale centrée donnée. Il place le Big Data en capacité stratégique commune multi-métier et multi-usage et en fait la clé de voûte du système d'information pour gérer la multi-modalité et l'interopérabilité entre l'ensemble des systèmes le composant : les systèmes opérants (e.g. Legacies, System-of-records), les systèmes informants (e.g. Business Intelligence, System-of-insights) et les systèmes d'interaction (e.g. Facing Apps, Customer Relationship Management Systems, Systems-of-engagement).

Notre offre Big Data s'adapte à tous les niveaux de maturité Big Data métier et SI. Elle s'étend du cadrage à l'implémentation, y compris la conduite du changement et la communication. Elle couvre la conception et mise en place de « Data Fabric » (Data Lake, Data Factory et Data Lab) internalisée ou externalisée, centralisée ou distribuée. Elle va jusqu'à l'activation de centre de services Big Data avec l'accompagnement à l'évolution des compétences internes si nécessaire. Elle s'appuie sur notre cadre méthodologique éprouvé et reconnu et nos nombreux accélérateurs Big Data, fruits de notre expérience unique sur le marché : smartfoundations™.

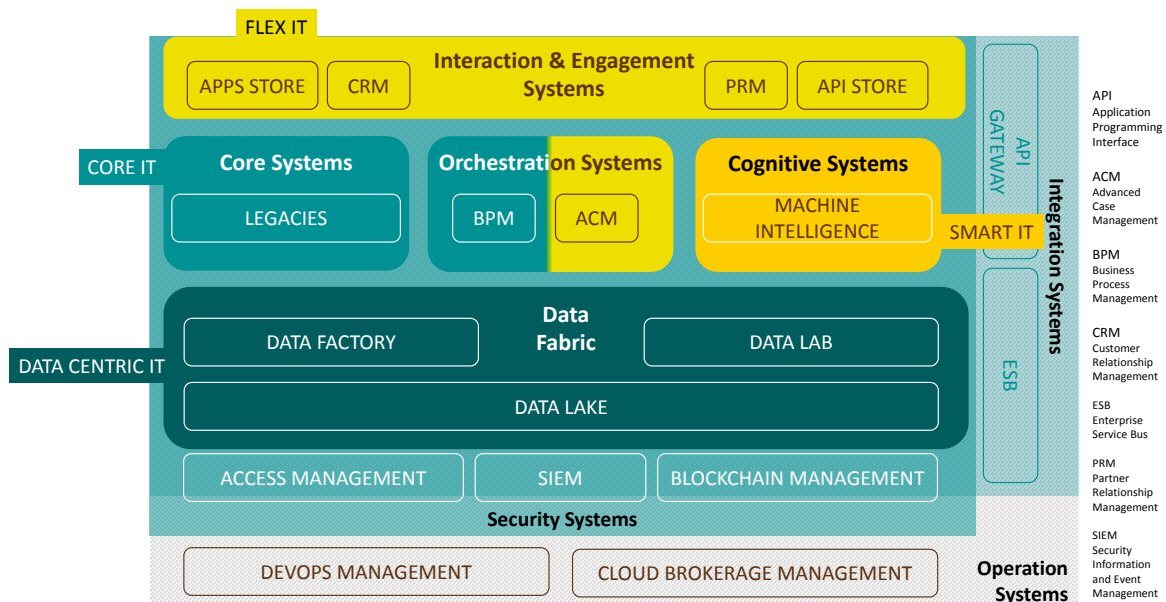
Notre offre Big Data a déjà séduit des acteurs leaders de leur secteur d'activité comme COVEA (Assurance), CREDIT AGRICOLE (Banque), CARREFOUR (Distribution), PMU (Paris sportif) ou encore AG2R LA MONDIALE (Santé & Prévoyance) pour ne citer que quelques-unes de nos références.

Nous accompagnons la plupart de ces clients depuis le début de leur aventure Big Data. Celle-ci commence en général par la réalisation d'un PoC (preuve-de-concept) pour démontrer, dans une logique d'architecture data-centrée, que les technologies Big Data sont les seules à pouvoir répondre aux nouveaux usages digitaux mais qu'elles apportent également une nouvelle réponse à des points de souffrance existants comme jamais auparavant. Par exemple capacité/vitesse de croisement de données, réduction des coûts des systèmes patrimoniaux (moins d'appels de services, sortie de fonctions non cœur), capacité d'analyse automatisée et de data science outillée et industrialisée, réduction drastique des temps de traitements décisionnels, capacité de convergence rapide en phase de fusion d'usines de gestion, etc.

Passée cette étape indispensable, nous co-construisons avec eux et sur l'ensemble des dimensions (infrastructure technique, socle logiciel, organisation et offre de services) leur Data Fabric en y intégrant souvent une capacité de Data Lab. Ce socle posé, nous réalisons systématiquement les premiers usages opérationnels jusqu'à leur mise en production. Parmi les usages adressés on peut citer : V360 temps réel, détection de fraude, détection du risque d'attrition, détection d'opportunités commerciales, gestion de stocks temps réel, audit de flux financier ou encore amélioration des performances de chaîne décisionnelle. Nous les accompagnons ensuite dans la conception et l'activation de l'organisation pour pérenniser leur dispositif industriel.

Quels que soient les déclencheurs à l'origine de leur initiative Big Data, tous ces clients partagent notre conviction. Le Big Data se pense dans une vision globale d'architecture data-centrée et la fondation Big Data qui la supporte doit dissocier les socles (technique, logiciel, données) des usages et pérenniser le tout via une organisation transverse (e.g. centre de services) garante de la cohérence entre les différents cycles de vie et de la prise en compte sur son périmètre de responsabilité des enjeux de gouvernance des données.

Bien que tous nos clients ont une histoire différente, ils ont tous pris la mesure que passer d'un PoC à un usage en production, qui plus est as-treint aux exigences du digital réclame beaucoup plus d'effort que le bat-tage médiatique ne le laisse entendre. Tous témoignent de la complétude et de la pertinence de notre vision et de notre capacité à la leur délivrer avec un haut niveau de qualité et d'engagement et ce, de façon différenciée et actionnable. Tous apprécient notre valeur fondamentale : mettre en œuvre ce que nous conseillons et concevons.



©Sentelis 2016 – smarTarchitecture 3.0 Canevas

**B I G**  
**DATA**

ARCHITECTURE  
LAKE  
FACTORY  
LAB

  
**sentelis**  
smart solutions for smarter enterprises.

Un jour

smartfoundations 

**VOUS Y VIENDREZ.**

[www.sentelis.com](http://www.sentelis.com)

31 rue du pont - 92200 Neuilly-sur-Seine - Tél : 01 75 61 34 00

BIG DATA Paris  
Consulting Sponsor  
Depuis sa création





## L'IRSN CHOISIT SINEQUA POUR VALORISER SON PATRIMOINE DE CONNAISSANCES ET SON SAVOIR-FAIRE SCIENTIFIQUE ET TECHNIQUE.

**Contexte :** L'Institut de Radioprotection et de sûreté Nucléaire (IRSN) est un Etablissement Public à caractère Industriel et Commercial (EPIC), de 1700 personnes, sous la tutelle conjointe de six ministères. L'IRSN est l'expert public en matière de recherche et d'expertise sur les risques nucléaires et radiologiques.

La connaissance est une valeur fondamentale dans le domaine d'évaluation et de la maîtrise des risques nucléaires et radiologiques. Une nouvelle fonction dédiée au management des connaissances au sein de l'IRSN a été créée en 2012. Elle s'est traduite par la nomination d'un Directeur du management des connaissances dont la mission est de doter l'institut d'une vision globale des connaissances clés à détenir à court, moyen et long terme, d'organiser la mise à disposition des connaissances et d'optimiser les dispositifs d'apprentissage et de partage ainsi que les bonnes pratiques managériales associées.

### Les enjeux pour l'IRSN : amélioration des connaissances et de l'efficacité des experts et des chercheurs

Les premières analyses du référentiel de connaissances menées dans le domaine de la sûreté ont mis en évidence la nécessité de structuration des connaissances en référentiel d'entreprise. Par ailleurs, il a été nécessaire de mettre en place le « portail des connaissances » : un outil permettant de retrouver facilement l'historique des sujets d'expertise traités et d'accéder aux méthodes et raisonnements utilisés tout en visualisant rapidement et directement les éléments de texte pertinents. L'objectif ? Améliorer les connaissances et par conséquent l'efficacité des collaborateurs par un accès simple et instantané aux expertises réalisées au cours du temps par l'Institut.

### Les challenges : retrouver et exploiter l'historique des sujets d'expertise

L'IRSN possédait, jusqu'à présent, un outil permettant une recherche sur les métadonnées offrant des résultats satisfaisants, sans toutefois permettre une recherche plus générale, sans limitation aux métadonnées. La nouvelle plateforme voulu par l'Institut se devait d'être accessible via une interface Web et permettre aux utilisateurs de réaliser une recherche en « texte libre ». Pour cela, il fallait que l'outil en question propose, non

seulement une liste de documents, mais également un accès direct aux zones des documents les plus pertinentes en fonction de la recherche. Une majeure partie des documents, principalement en français et en anglais, formalisant l'expertise de l'IRSN a été numérisée dans des bases de données internes. Afin de faciliter le travail des collaborateurs, il est important qu'il soit possible d'effectuer une recherche simple et efficace au sein même de ces bases de données.

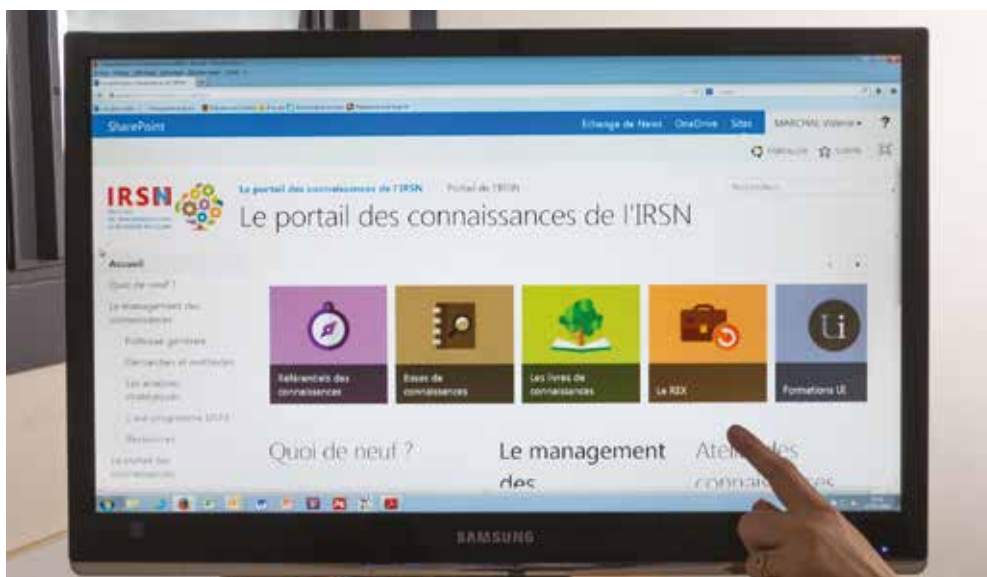
**La plateforme de Cognitive Search de Sinequa :** Le portail de connaissances s'appuyant sur Sinequa et baptisé ASK (Always Seek for Knowledge) a été retenu par l'IRSN après une étude comparative menée sur plusieurs solutions de recherche, suivie de campagnes de tests et d'une sélection finale fondée sur la pertinence, la contextualisation, la rapidité et la convivialité de l'interface.

La plus-value de l'outil mis en place avec Sinequa réside dans sa capacité à rechercher les documents sur la base d'une question en langage naturel et d'accéder directement aux passages les plus pertinents des avis et rapports, permettant ainsi de retrouver tout l'historique des analyses et raisonnements sur un thème donné.

La technologie Sinequa permet également à l'IRSN, confronté à une transition générationnelle, de s'adapter facilement à la jeune génération. C'est un outil simple à adopter et à s'approprier pour accéder aux connaissances historiques de l'institut. Compatible avec les terminaux mobiles, la plateforme de Sinequa incite donc à la mobilité, une caractéristique particulièrement appréciée de ses utilisateurs.

« Nous avons besoin d'un outil capable de réaliser une recherche sur plus de 40 années de production de connaissances dans l'expertise de sûreté nucléaire. Grâce à Sinequa, nous allons pouvoir retrouver facilement ces connaissances clés et ainsi faire valoir notre expertise dans ce domaine. » - Martial Jorel, Directeur du Management des connaissances, IRSN.

**Les résultats :** La plateforme mise en place par Sinequa est encore en cours de déploiement. Des extensions de domaines de couverture sont en cours (recherche en sûreté ...) ou programmées (radioprotection, environnement...). Néanmoins à ce stade, il remplit parfaitement sa mission : il offre une amélioration des connaissances et de l'efficacité des experts par un accès aux études réalisées au cours du temps par l'Institut.



## COGNITIVE SEARCH & POWERFUL ANALYTICS



### The Next Generation Insight Platform

- ✓ Instant Insight Across All Data Sources
- ✓ Advanced Natural Language Processing
- ✓ Deep Analytics of Big Data with Machine Learning
- ✓ Powerful Search-Based Applications



## RETOURS D'EXPERIENCES PROJETS DE MACHINE LEARNING

### ANALYSE SÉMANTIQUE POUR CONSTITUTION AUTOMATIQUE DE PANIERS PERSONNALISÉS

L'optimisation des ventes sur un site marchand passe par la fluidité du parcours client. En effet, la rapidité du processus de commande est un des facteurs essentiels de fréquentation des sites marchands généralistes, notamment alimentaires. Au-delà des fonctions de base comme la liste de courses habituelles, les fonctionnalités avancées permettant de faciliter la constitution de panier sont reconnues comme des leviers d'augmentation des ventes. Un usage possible consiste à proposer aux clients de choisir une recette de cuisine qu'ils souhaitent préparer, et d'ajouter automatiquement la liste des produits nécessaires dans son panier d'achats.

La technologie de SKAPANÉ utilise un réseau de neurones capable de construire un modèle sémantique sur la base d'un corpus de recettes de cuisine. L'algorithme mis en œuvre est ensuite capable d'associer des mots ayant le même sens même sans aucune racine lexicale commune. La portée de cette approche est importante, car il s'agit d'une technique d'apprentissage non-supervisé, ne nécessitant pas l'intervention humaine. Au-delà des aspects technologiques, la solution prend en compte des aspects métiers liés au commerce en ligne :

- Produits de substitution en cas de rupture de stock ou d'absence du produit dans le catalogue,
- Priorité est donnée aux produits à plus forte marge,
- Gestion des quantités afin de constituer un panier complet,
- Personnalisation des produits en fonction des habitudes de consommation du client, en intégrant son historique de consommation : marque de prédilection ou types de produits (bio par exemple).

La performance de calcul autorise le déploiement complètement on-line, permettant au client de choisir une recette en ligne, son panier étant automatiquement constitué. Il est également possible de « deviner » les recettes possibles sur la base d'un panier en cours de constitution et de proposer des produits complémentaires pour compléter la recette, ou des produits d'accompagnement (vins associés ou « petite touche » supplémentaire pour une recette).

La qualité de la solution est assurée par un contrôle des résultats sur la base du catalogue du e-commerçant.

Ces solutions font partie des outils facilitant le parcours d'achat des clients, en incluant la personnalisation client par client, et pas uniquement par catégories de clients. L'enjeu est la fidélité des clients, la mise en avant de produits et l'up-sell de produits compagnons. L'infrastructure logicielle nécessaire est complètement standard : utilisation de Spark (MLlib) et s'intègre avec toutes les solutions de e-commerce via des APIs de type REST (WebServices). Cette technologie est disponible sous forme de logiciel installé sur le site ou en mode SaaS.

### SÉCURISATION DU PAIEMENT EN UN CLICK SUR LES SITES MARCHANDS GRÂCE À L'APPRENTISSAGE AUTOMATIQUE EN TEMPS RÉEL

Le taux de fraude moyen sur les sites marchands français impose l'utilisation de solutions de sécurité renforcée. En effet, avec un chiffre au-dessus de 0,20% des transactions, le préjudice financier potentiel est important. Les solutions d'authentification forte répondent très bien à la problématique de sécurisation. Comme toute technique de sécurité, elle amène néanmoins des contraintes et n'est pas capable de détecter tous les types de fraude. En effet, la nécessité de confirmer un paiement par un code envoyé par SMS ou message sur smartphone, ou l'utilisation de boîtiers personnels de sécurité, voire de matrices de codes, est connue pour provoquer des abandons de panier : rallongement du temps de commande, non disponibilité du dispositif de sécurisation, ... D'autre part, sur certains types de paiement (plusieurs fois sans frais), la sécurisation ne couvre que la premier des paiements. Il est donc intéressant de compléter la sécurité du paiement en tant que tel par un outil de détection du risque de fraude à la sortie du tunnel de commande. Cette approche permet de choisir le type de paiement proposé au client (paiement sans autorisation, autorisation, autorisation 3DS, contrôle manuel de la commande, contrôle avec contact client, refus, ...) à la validation du panier.

La technologie de SKAPANÉ s'appuie sur un des algorithmes les plus puissants dans ces domaines : les forêts aléatoires (Random Forests). Sur la base de l'historique des commandes, un algorithme de détection des comportements et des caractéristiques des achats est calculé, permettant d'estimer en temps réel un niveau de risque de fraude pour la commande en cours. La solution est complétée par un moteur de règles métiers permettant aux gestionnaires de sites et au gestionnaires de fraude de contrôler le comportement dynamique du système en y adjoignant des règles métiers spécifiques (contrôle systématique au-dessus d'un certain montant, ou pour certains types de produits, modulation du risque par la marge du panier en cours, ...) ainsi que d'indicateurs temps réel des choix de routage effectués par le système.

L'amélioration de qualification de transactions potentiellement frauduleuses permet au système de vente de faciliter l'expérience utilisateur des clients en identifiant précocement les transactions à risques et ainsi simplifier le processus de paiement des clients à faible risque en allant jusqu'au paiement en 1 click pour la plus grande proportion possible des transactions.

L'infrastructure logicielle nécessaire est complètement standard : utilisation de Hadoop, Spark (MLlib) d'une base de données NoSQL haute performance (Aerospike) et s'intègre avec toutes les solutions de e-commerce via des APIs de type REST (WebServices). Cette technologie est disponible sous forme de logiciel installée sur le site ou en mode SaaS.





### QUI EST SKAPÁNĚ ?

SKAPÁNĚ est une start-up innovante spécialisée dans les traitements big data en temps réel et le Machine Learning, accélérée dès son lancement début 2015 au sein d'Euratechnologies à Lille. Notre vocation est d'apporter à nos clients la haute technologie de la prédiction, de la recommandation produit, de la classification clients, de la lutte contre la fraude, de l'analyse sémantique et du marketing relationnel. Une de nos différenciations réside en notre capacité à intégrer ces technologies de pointe en temps réel au sein des chaînes transactionnelles quand le marché le propose principalement sous forme de rapports d'analyse ou au mieux d'intégration batch. Notre démarche consiste à rester pragmatique et à mesurer des résultats à chaque étape : analyse stratégique, formation, mise en œuvre de pilotes, de projets, exploitation ... Nous proposons un accompagnement agile de bout en bout dans la mise en œuvre d'algorithmes sophistiqués dont la puissance permet de traiter des énormes volumes de données disponibles mais encore inexploitées : données internes, externes, open data...

### QUELS SONT LES TYPES DE PROJETS QUE VOUS RÉALISEZ ?

Actuellement nous menons des projets très variés. Nous formons beaucoup de data scientists au traitement des données volumineuses avec Hadoop & Spark par exemple chez Axa, Cofidis, MonaBanq ou Oney Banque Accord. Nos formations sont très appréciées car nous y intégrons beaucoup de pratique, les stagiaires manipulent leurs propres données et nous leur mettons ainsi le pied à l'étrier pour intégrer un projet de datascience. A la suite des formations, nous continuons d'accompagner ces clients sur leurs projets. Nous formons ou plutôt coachons également des membres de comités de direction et des managers qui ont besoin d'être accompagnés à la transformation numérique sur la dimension de la gestion de la donnée, et notamment de la donnée volumineuse ou non-structurée. Bien évidemment nous menons des projets de Machine Learning pour le compte de nos clients. A titre d'exemple, nous réalisons pour C Discount un projet de lutte en contre la fraude en Apprentissage Automatique : l'analyse du comportement des clients sur le site via des algorithmes de Machine Learning mis en place par SKAPÁNĚ permet de définir un score de risque de fraude et ainsi d'autoriser le paiement en un click pour les bons clients, augmentant ainsi le taux de transformation. Pour Auchan Retail data, nous réalisons un projet de constitution automatique de paniers basé sur de l'analyse sémantique de recettes de cuisine.

### PARLEZ-NOUS DE VOTRE PLATE-FORME CLOUD

Effectivement, nous mettons également en œuvre des projets en mode Cloud, c'est par exemple ce que nous réalisons dans le mode de la connaissance clients. Nous opérons pour le compte d'un de nos clients du monde du Marketing Relationnel une plate-forme de comptage et d'extraction à des fins de télémarketing. Dans ce cas SKAPÁNĚ développe et d'exploite la base de données contenant plus de 40 millions d'individus, et un peu moins de 100 millions de contacts [adresse postale / e-mail/ tel fixe, tel mobile]. La plate-forme permet d'extraire des listes de contacts sur la base de critères simples (localisation géographique par exemple), complexes (profil client sur la base de ces achats / prédiction d'intentionnistes / détection d'événements de vie ...) ou issus de l'Open Data (données du recensement de la population par exemple). Sur cette plate-forme, nous réalisons aussi des projets de connaissance clients pour Cofidis : analyse de la navigation des clients et prospects sur le site web, analyse des messages reçus et tri automatique par nature, urgence... L'offre en mode Cloud est une offre que nous souhaitons développer car elle permet à nos clients de lancer des pilotes, des expérimentations .... Afin de tester des usages et de démontrer les apports de ces technologies dans leur business. C'est une approche très pertinente dans notre domaine, où l'agilité et la rapidité sont clés.

### ET DANS LE CAS DU MODE CLOUD, OÙ SONT LES DONNÉES ?

Bien sûr, notre plate-forme, et donc les données, est localisée dans un Datacenter extrêmement sécurisé, bénéficiant des certifications PCI DSS du monde du paiement et HADS du monde de la santé, situé dans les Hauts de France. Il n'est pas question pour nous de mettre les données de nos clients ailleurs qu'en France. Nous veillons également à ce que tous les aspects légaux et juridiques soient respectés : déclarations CNIL, protection des données personnelles...

### QUELLES SONT LES TECHNOLOGIES DE RÉFÉRENCE CHEZ SKAPÁNĚ ?

La vague de développement de logiciels d'infrastructure Open Source est particulièrement réussie sur le monde du Big Data et du Machine Learning. Plusieurs solutions de référence de grande qualité sont disponibles. Nous avons choisi Hadoop et Spark pour le stockage distribué sécurisé et les traitements distribués. Nous utilisons les bases de données NoSQL du marché : HBase, Cassandra pour les bases orientées colonnes ou Aerospike pour les besoins de performance temps réel. Spark, particulièrement depuis la version 2.0, propose une librairie très complète sur les algorithmes de Machine Learning modernes. Nos plateformes en production nous ont permis de démontrer la qualité, la production sécurisée et le coût de possession attractif de ces solutions.

### QUELLE EST VOTRE VISION POUR SKAPÁNĚ ?

Nous souhaitons aider à la mise en œuvre de solutions de traitements de données et de Machine Learning en France et en Europe. En effet, il s'agit là d'un des piliers de la transformation numérique, qui nous semble encore peu développé sur le vieux continent. Nous espérons contribuer à la formation d'ingénieurs experts, de Data scientists et à l'introduction de ces technologies dans tous types de projets d'entreprises, car elles ont démontré leur capacité à amener certaines des start-ups de la Silicon Valley à devenir des géantes aujourd'hui.

Sandra GOMES

Directrice Commerciale et Marketing

+33(0)1 73 00 55 00

contact@softcomputing.com

55 quai de Grenelle 75015 Paris France



www.softcomputing.com

soft.computing

## SOFT COMPUTING ACCOMPAGNE SES CLIENTS POUR EXPLOITER TOUT LE POTENTIEL DE LA DATA, CRÉER DES EXPÉRIENCES CLIENT SANS COUTURE ET DÉMULTIPLIER LA PERFORMANCE DU MARKETING DIGITAL.

### BIG DATA : L'ÈRE DU DATA-DRIVEN

#### Big Data : nouvelle frontière

Les volumes de données produites augmentent de manière exponentielle avec l'explosion des objets connectés et des applications sociales et mobiles. L'analyse intelligente de ces énormes gisements de données transforme notre société. Le marketing, en particulier, devient la clé de voûte de la stratégie des entreprises puisqu'il dispose des données et de la connaissance des clients, des produits et des canaux.

Parallèlement à l'explosion des volumes, la puissance de calcul, les capacités de stockage et le cloud révolutionnent le champ du possible.

#### Passer au data-driven marketing : de la maquette à l'exploitation Big Data

Soft Computing aide ses clients à tirer profit de ces nouvelles possibilités pour passer à un marketing temps réel et guidé par les données.

Identifier des parcours clients sur Internet, détecter des tendances en écoutant les réseaux sociaux, automatiser les enchères pour l'achat d'espaces publicitaires sur le Web, mettre en place des recommandations auto-apprenants, personnaliser en temps réel un site en fonction d'informations CRM ou de visites précédentes, telles sont quelques applications concrètes au marketing de nos savoir-faire en matière de Big Data couplé avec les traditionnels Datawarehouses. Pour se faire, nous intervenons à la fois sur les dimensions fonctionnelles et techniques, depuis la réflexion jusqu'à l'exécution :

- Sensibilisation à la valeur des données ;
- Identification des cas d'usages et de leur ROI potentiel ;
- Réalisation de Proof of Concept Big Data ;
- Couplage entre CRM et Big Data ;
- Conception d'architectures Big Data et Datawarehouse ;
- Implémentation de solutions Big Data ;
- Gestion technique des solutions Big Data ;
- Gestion opérationnelle de processus marketing Big Data : écoute des réseaux sociaux, moteurs de recommandations, optimisation des parcours client, personnalisation des contenus, DMP, RTB...

### DATALAKE READY® : UNE OFFRE BIG DATA PRÊTE À L'EMPLOI FORGÉE SUR NOS RETOURS D'EXPÉRIENCE

#### Votre plate-forme pour apprendre, tester ou débiter le déploiement

La valeur ajoutée du Big Data se trouve dans les usages analytiques et non dans les infrastructures qui ne sont que des moyens. Nous avons construit l'offre Datalake Ready®, pour offrir à nos clients un environnement Big Data prêt à l'emploi (As A Service). C'est une solution simple et sécurisée, sous forme d'abonnement mensuel économique, sans investissement, pour être rapidement opérationnelle.

Concrètement c'est un Datalake qui permet de stocker les données structurées ou non structurées (transactions, Weblog, Logserveur, Digital In-Store, tweets, post). Au-delà d'une solution technique, nous proposons un dispositif complet qui comprend la mise à disposition de l'infrastructure, mais aussi les services pour opérer le Big Data (les uses cases) en fonction des besoins et des attentes des clients.

Elle permet de ne pas avoir à se lancer dans un projet long, couteux et parfois disproportionné, de franchir le cap du Big Data pour faire face à des problématiques métiers. Avec Datalake Ready® il est alors possible de :

- Disposer d'un environnement Big Data prêt à l'emploi ;
- Explorer de nouveaux gisements de données ;
- Collecter des données volumineuses pour une utilisation immédiate ou future ;
- Acquérir des compétences Big Data sans engager d'investissements (faire un POC Big Data) ;
- Tester des outils d'analyse sur du Big Data.

#### Notre proposition Datalake Ready® comprend dès le « Starter Pack » :

- Une plate-forme Big Data élastique, administrée et sécurisée dans un Data Center en France ;
- Le stockage, l'hébergement et la supervision des flux ;
- L'ensemble des composants de la distribution Cloudera ;
- Une « tranche » de 10 to de données sécurisées ;
- La garantie dès l'amont des conditions de réversibilité pour assurer la possibilité de ré-internaliser au rythme souhaité tout ou partie de la solution ;
- Ce pack est personnalisable.

De nombreux services optionnels sont disponibles, tels que la mise à disposition des outils analytiques, BI ou des services métiers avec un vivier d'experts en Data Science et Big Data pour vous accompagner dans vos premiers cas d'usages (80 data scientists, experts métiers usages, data managers, architectes).

Des missions de cadrage sont aussi possibles pour un accompagnement dans la réflexion sur les use case et leurs ROI.

### UNE OFFRE UNIQUE MAÎTRISÉE DANS LE TEMPS ET ACCESSIBLE À TOUS LES BUDGETS

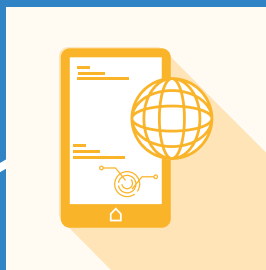
Datalake Ready® est une solution complète pour servir les Directions métiers & DSI en les affranchissant des contraintes technologiques de mise en œuvre, d'exploitation et de sécurité.

Avec Datalake Ready®, disposez d'un environnement Big Data en quelques jours, collectez et conservez toutes vos données, découvrez les opportunités et les gisements de valeurs non exploitées jusqu'à présent, dotez-vous des outils analytiques et marketing nécessaires pour rester dans la course, sans engager des projets longs et couteux.





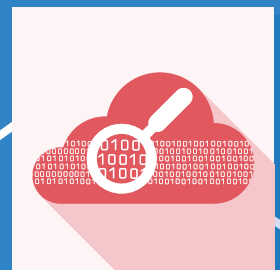
Exploiter tout le potentiel de la data.  
Créer des expériences Client sans couture.  
Démultiplier la performance du marketing digital.



Digital



Marketing



Big Data

## La confiance des leaders



## SOPRA STERIA

**Pierre ZALDIVAR**

Directeur des Pratiques Digitales  
pierre.zaldivar@soprasteria.com

**Manon LINEL**

Responsable Marketing BigData  
manon.linel@soprasteria.com

9 bis rue de Presbourg - 75116 Paris France



www.soprasteria.com



**sopra steria**

Composante majeure du Digital Transformation Office de Sopra Steria, forte de plus de 1000 collaborateurs en Europe, la Practice Analytics & Big Data :

- Fédère et anime les équipes d'experts et architectes sur les projets Big Data, Data Science, MDM et décisionnels du Groupe ;
- Crée les parcours de formations pour nos collaborateurs (250 formés par an sur Hadoop, Datascience, Architectures Big Data) ;
- Capitalise les bonnes pratiques (gouvernance des données des DataLake intelligents, industrialisation des projets Hadoop, architecture à l'état de l'art, ...) ;
- Met à disposition et administre les environnements Big Data Ready pour nos clients.

L'offre Analytics et Big Data de Sopra Steria se déploie principalement sur les thématiques suivantes :

- Valorisation des données clients (parcours omni-canal, données Internet, données d'usage des produits/services, ...) pour améliorer la relation client,
- Analyse des logs des capteurs et de l'IoT pour la maintenance préventive,
- Analyse sémantique des données non structurées (email, voix du client, CV, ...) pour en tirer le meilleur parti,
- Passage du mode réactif au mode prédictif à l'aide des nouveaux outils algorithmiques et analytiques.
- Exploitation des nouveaux modes de visualisation des données pour mieux les valoriser,
- Intégration des nouvelles technologies Hadoop, Appliances, In Memory... pour servir les enjeux et cas d'usage métier.

Parmi les réalisations récentes de Sopra Steria en Europe nous pouvons citer :

### SOPRA STERIA DÉLIVRE POUR LA FCA, LA PLUS GRANDE PLATE-FORME DIGITALE DE TRAITEMENT DES DONNÉES DE MARCHÉ DU ROYAUME-UNI

Dans ce cadre, la nouvelle plate-forme de traitement des données de marché est un support à la régulation actuelle des entreprises mise en place par la FCA. La solution mise en œuvre par Sopra Steria permettra également de traiter les directives actualisées de la réglementation concernant les marchés d'instruments financiers (MiFID II).

Un entrepôt gigantesque contenant des milliards de transactions avec une profondeur d'historique de 6 années minimum constitue le cœur de la solution. La FCA l'exploitera à des fins de transparence et de traçabilité. La plateforme a été également conçue pour être partagée afin de profiter à d'autres organismes. Technologiquement la plateforme exploite les solutions Hadoop, Amazon et Axway pour assurer la réception et la délivrance des transactions et DataStax pour assurer la conformité aux règles MiFID II.



**Gareth Lewis**  
CIO de la FCA

«Disposer d'une solution de traitement des données de marché fiable, robuste, évolutive est un impératif dans notre rôle de régulateur »

### Pascal Lambert Responsable SIV DDTN-DV Distribution Validation

« Les équipes de Sopra Steria nous apportent les compétences nécessaires pour réussir la mise en œuvre des nouvelles technologies associées au Big Data, afin de servir les enjeux essentiels de la stratégie digitale de SNCF et TRANSILIEN »

### SOPRA STERIA DÉVELOPPE LE DATALAKE SNCF TRANSILIEN POUR SERVIR LES ENJEUX DIGITAUX

SNCF a pour ambition de rendre simple et fluide le transport des personnes et des marchandises en France et dans le monde. Au sein de SNCF Mobilités, TRANSILIEN est l'offre de transport de voyageurs en Ile-de-France. TRANSILIEN c'est 3,2 millions de voyageurs par jour soit 70% de voyageurs SNCF, +30% de voyageurs en 10 ans, 6200 trains par jour sur 10% du réseau ferré français.

Au travers de sa stratégie digitale, SNCF souhaite se transformer d'un service ferroviaire en acteur de solutions de mobilité innovantes, en étant à la pointe de la technologie dans la relation clients, la pratique professionnelle de l'ensemble des salariés ou encore la relation avec les startups.

Le Big Data est un nouvel outil pour TRANSILIEN afin d'améliorer la performance, l'efficacité et la qualité de service du transport en Ile de France vis-à-vis de ses clients et du STIF (Syndicat des Transports en Ile de France, autorité organisatrice des transports de la région Ile de France).

En lien étroit avec la Big Data Fab de la Direction du Digital SNCF, TRANSILIEN a démarré son initiative Big Data en 2015 autour d'une infrastructure Hadoop. TRANSILIEN est accompagné par Sopra Steria depuis le démarrage de ces initiatives dans une approche Prototype et désormais dans leur mise en œuvre industrielle : Sopra Steria appuie les équipes projets TRANSILIEN à la mise en œuvre du Big Data :

- Mise en place d'une architecture répondant aux exigences métiers et aux exigences de sécurité de SNCF ;
- Réalisation en mode agile de mécanismes industriels d'ingestion de données hétérogènes (ex : données de validations, données de comptage, données de connaissance du client, ...) ;
- Développement de cas d'usage à forte valeur ajoutée (ex: outil de recherche pour la Sureté ferroviaire).

### SOPRA STERIA AMÉLIORE LA DÉTECTION DE LA FRAUDE FISCALE EN ESPAGNE

Les équipes Sopra Steria développent et implémentent des algorithmes de Machine Learning et d'Intelligence Artificielle pour améliorer la détection de la fraude fiscale. Ces nouvelles techniques permettent de segmenter les classes de fraudeurs et de détecter la fraude organisée en identifiant les intermédiaires entre des classes de fraudeurs apparemment sans relation, via l'utilisation des données issues des réseaux sociaux.

# Sopra Steria & **le big data**



+ 38 000  
COLLABORATEURS

+ 3,6 MDS €  
DE CA PRO FORMA 2015

+ 20 PAYS  
EN EUROPE ET DANS LE MONDE

+ 45 ANS  
D'EXPERTISE

## DES SECTEURS CIBLÉS

Banque & Assurance -  
Secteur Public - Aéronautique,  
Défense & Sécurité - Transport -  
Télécoms & Médias - Énergie -  
Distribution

**Sopra Steria, leader européen de la transformation numérique, propose l'un des portefeuilles d'offres les plus complets du marché :** conseil, intégration de systèmes, édition de solutions métier et Business Process Services.

Sopra Steria apporte une réponse globale aux enjeux de développement et de compétitivité des grandes entreprises et organisations, notamment en relevant les défis du Big Data :

- comprendre et analyser les comportements possibles avec l'internet des objets, les réseaux sociaux...
- identifier de nouveaux usages à valeur ajoutée,
- vous différencier par de nouveaux services.

[www.soprasteria.com](http://www.soprasteria.com)





**HISTORIQUE**

- 2006 Splunk 1, Splunk 2; 150 clients
- 2007 Splunk 3 ; 350 clients
- 2008 Expansion APAC et EMEA ; 900 clients
- 2009 Splunk 4 ; 1 000 clients
- 2010 Splunk 4.1 ; Première conférence internationale des utilisateurs ; 2 000 clients
- 2011 Splunk 4.2; Ouverture du département de recherche et développement de Seattle ; plus de 3 000 clients
- 2012 Splunk Enterprise 5 ; plus de 4 800 clients
- 2013 Splunk Enterprise 6 ; Splunk Cloud; Hunk ; plus de 7 000 clients
- 2014 Splunk Enterprise 6.2, Splunk MINT ; Application Splunk for Stream ; plus de 9 000 clients
- 2015 Splunk Enterprise 6.3 ; Splunk ITSI ; Splunk UBA ; plus de 10 000 clients
- 2016 Splunk Enterprise 6.5 ; plus de 12 000 clients

**PRINCIPAUX SECTEURS D'ACTIVITÉS**

- Communications
- Énergie et Infrastructures
- Finances
- Santé
- Enseignement supérieur
- Industrie
- Organisation à but non lucratif
- Services en ligne
- Secteur public
- Distribution

**SPLUNK EN CHIFFRES**

Nombre de clients = Plus de 12 000  
 Applications Splunk = Plus de 1 000  
 Clients du Fortune 100 = Plus de 80  
 Siège = San Francisco  
 Employés = Plus de 2 700 dans le monde  
 Cotation= NASDAQ – SPLK - investors.splunk.com

**RENDRE LES DONNÉES MACHINE ACCESSIBLES, UTILES ET EXPLOITABLES PAR TOUS.**

Qu'est-ce que Splunk ? Grâce à Splunk, vos données machine prennent du sens, que ce soit sur le plan informatique, de la sécurité, commercial, ou métier. Les solutions Splunk permettent à votre département IT et aussi à toute votre entreprise d'obtenir une meilleure visibilité de vos activités et de découvrir de nouvelles perspectives. Splunk a été fondé sur la base d'une vision radicalement innovante : rendre les données machine accessibles, utiles et exploitables par tous.

**DONNÉES MACHINE**

Les données machines sont l'un des segments les plus porteurs et les plus complexes des « big data ». Splunk transforme les données machines en renseignements précieux quel que soit votre secteur. C'est ce que nous appelons l'Intelligence Opérationnelle.

**LA PLATEFORME SPLUNK INTELLIGENCE OPÉRATIONNELLE**

L'Intelligence Opérationnelle vous offre une vision en temps réel de ce qui se passe sur l'ensemble de vos systèmes informatiques et infrastructures technologiques, à n que vous puissiez prendre des décisions éclairées, pouvant bénéficier à l'ensemble de l'entreprise.

**NOS SOLUTIONS**

Infrastructure et opérations informatiques. Détectez, analysez et visualisez les problèmes de réseau, de serveur, de stockage, de virtualisation ou d'infrastructure cloud, et mettez-les en corrélation avec les problèmes utilisateurs ou applicatifs. Vous pouvez ainsi profiter d'une surveillance en temps réel, d'alertes préventives et d'une visibilité sur la santé des services informatiques de tous les niveaux de votre organisation.

**GESTION DES APPLICATIONS**

Bénéficiez d'une vision en temps réel de la pile applicative dans son intégralité - Splunk fournit des vues d'ensemble sur la performance, les transactions et les activités d'utilisateur. Avec les logiciels Splunk, les DevOps publient plus fréquemment de nouvelles versions, les équipes opérationnelles réduisent leur temps moyen de résolution, et les équipes d'ingénierie optimisent la qualité, les performances et le coût des applications.

**SÉCURITÉ ET CONFORMITÉ**

Grâce à la surveillance des menaces de sécurité, l'analyse des historiques et la visualisation des ensembles de données massifs, bénéficiez d'une meilleure intelligence sécuritaire, tant pour les menaces connues qu'inconnues. L'exploration en temps réel des données d'incidents étant grandement facilitée, il est possible de mener des enquêtes complètes, d'assurer une protection préventive, et de créer des rapports ad hoc en quelques minutes.

**BUSINESS ANALYTICS**

Les données machines renferment de précieux renseignements commerciaux pertinents et utiles pouvant conférer un avantage compétitif considérable. Bénéficiez d'une visibilité sur les processus métier complexes, le comportement des clients, l'utilisation des produits et les campagnes marketing digitales, le tout en temps réel.

**INTERNET DES OBJETS ET DONNÉES INDUSTRIELLES**

Exploitez les données générées par les appareils connectés, les systèmes de contrôle, les détecteurs, les systèmes SCADA et autres. Surveillez les opérations, analysez les usages et intégrez ces renseignements pour produire une image globale de vos opérations métier.

**NOS CLIENTS**

Plus de 12 000 clients dans 110 pays utilisent Splunk pour améliorer leurs niveaux de service, réduire les coûts opérationnels et les risques de sécurité, favoriser la conformité et créer de nouveaux produits et de nouvelles offres de service. Grâce à Splunk, elles peuvent dépasser les modes traditionnels de stockage de données, élever leur Intelligence Opérationnelle à un autre niveau, et obtenir des perspectives qui profitent à toute l'entreprise.

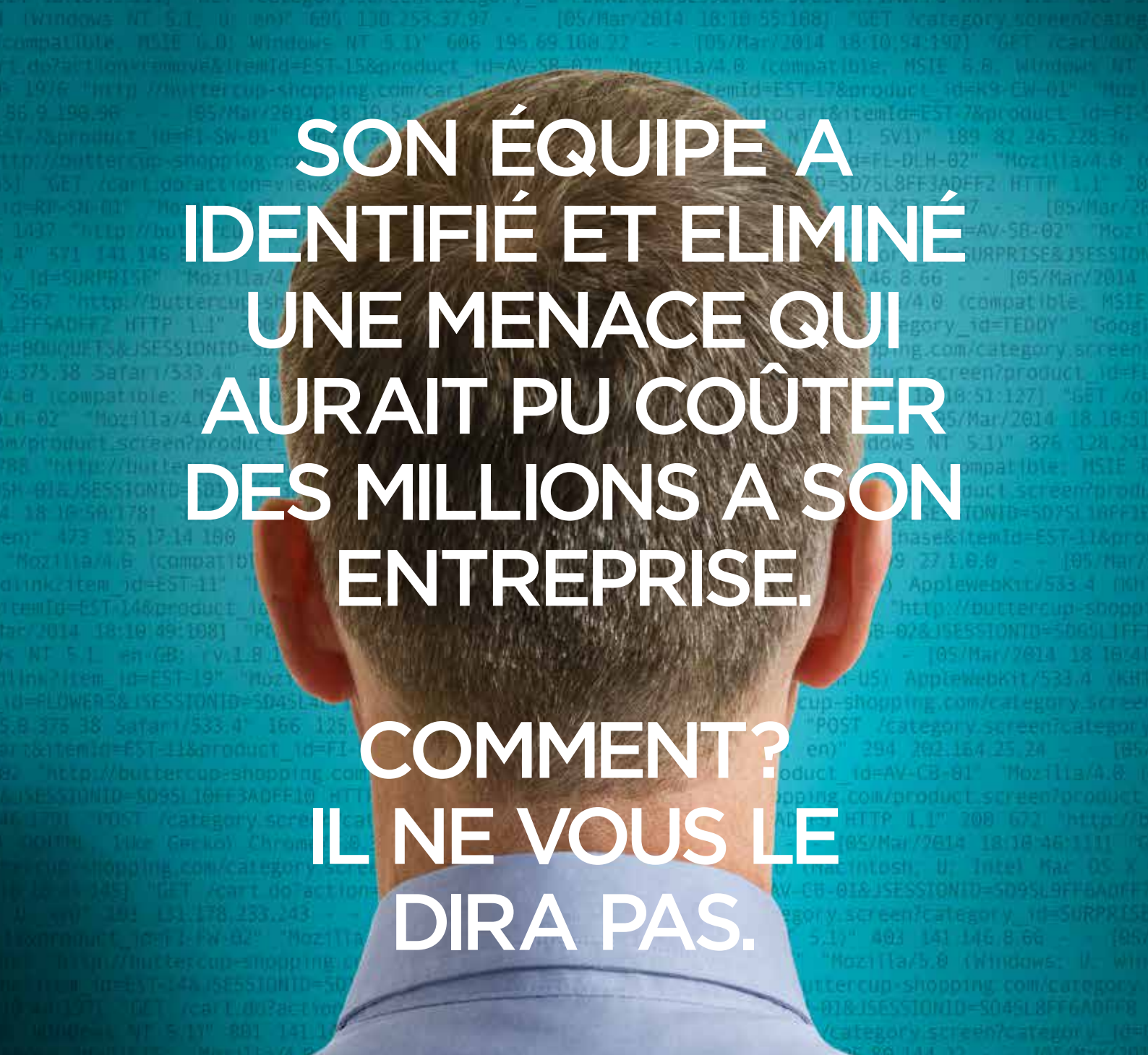
**Solutions Premium Splunk**

**Un riche écosystème d'applications et d'extensions**

splunk>enterprise

splunk>cloud

**splunk>** Plate-forme d'Intelligence Opérationnelle



**SON ÉQUIPE A  
IDENTIFIÉ ET ELIMINÉ  
UNE MENACE QUI  
AURAIT PU COÛTER  
DES MILLIONS A SON  
ENTREPRISE.**

**COMMENT?  
IL NE VOUS LE  
DIRA PAS.**

**Splunk® donne aux équipes de sécurité une visibilité sur l'ensemble de l'infrastructure pour qu'elles puissent rapidement détecter et contenir les menaces qui pèsent sur leur entreprise. Plus de 80 des entreprises du « Fortune 100 » utilisent Splunk pour être plus compétitifs et protéger leurs activités, mais peu d'entre elles vous le diront.**

**Demandez-vous**

**▶ Que pouvez-vous faire avec Splunk? Découvrez-le sur [fr.splunk.com](http://fr.splunk.com)**

**splunk®** listen to your data®

**Robert BOUNHENG**

Directeur du développement

+33(0)1 55 28 89 40

contact@synaltic.fr

24 rue de l'Église 94300 Vincennes France



www.synaltic.fr



CONTACT

## SYNALTIC MET À DISPOSITION UN OUTIL GÉO-DÉCISIONNEL PERSONNALISÉ ET PERFORMANT POUR BOUYGUES IMMOBILIER : MYGEOMARKET.

Bouygues Immobilier, filiale du Groupe Bouygues, développe en France, en Europe et au Maroc des projets de logements, d'immeubles de bureaux ou de commerces et aménage des quartiers durables. Publicis Média, nouvelle entité de Publicis, est l'agence média et conseil en communication locale et digitale de Bouygues Immobilier. Synaltic est spécialiste du data management, expert en open source, en SIG et en développement d'applications à façon et de plateformes data.

Le projet a été proposé à Bouygues Immobilier par Marien Lavoit, directeur commercial au sein de Publicis Media, à partir d'une conviction simple et largement partagée chez Bouygues Immobilier : une analyse précise de la géographie des ventes de Bouygues Immobilier permettrait d'améliorer significativement la précision et la qualité des recommandations de communication locale et donc in fine d'améliorer le ROI des actions media.

Ce dernier, également co-enseignant aux côtés de Meriadeg Jaouen, expert cartographie et SIG de Synaltic, au sein du Master Géomatique G2M de Paris 8, se tourne alors naturellement vers Synaltic. « Je cherchais un partenaire suffisamment agile pour travailler en design thinking avec une cellule marketing, ayant une connaissance poussée des systèmes d'information géographique et proposant des performances suffisantes pour une expérience utilisateur optimale ».

Fort de son expérience dans la mise en place de Data Platform, Synaltic a accepté le challenge et développé, en quelques mois et en collaboration avec les équipes métiers de Bouygues Immobilier, MyGeoMarket, une application géo-décisionnelle, simple d'utilisation et intuitive.

L'outil permet de sélectionner une zone géographique précise et d'y faire apparaître instantanément un certain nombre de critères issus des données ouvertes de l'INSEE (lieu, CSP des habitants, tranches d'âge, composition des ménages, etc.) et de les croiser aux données propres à Bouygues Immobilier, extraites de son système d'information (prospects/réservataires), ainsi qu'à des données consolidées fournies par Publicis Media.

## Données géographiques, open data, CRM, données consolidées, toutes dans le data lake !

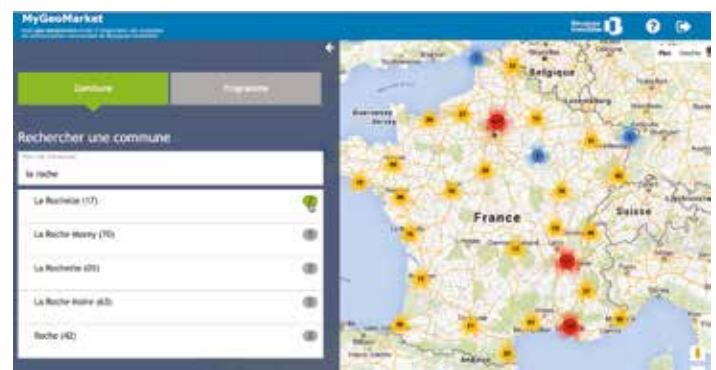
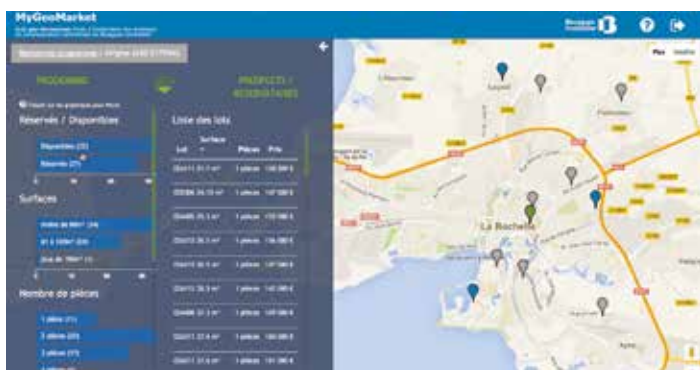
MyGeoMarket repose sur une base de données PostgreSQL – choisie pour sa robustesse et sa puissance – ainsi que son extension de traitement de données géographiques PostGIS et un backend Node.JS. Du côté de l'interface utilisateur, AngularJS et l'API Google Maps lui confèrent une simplicité d'utilisation et un rendu visuel optimal.

L'utilisateur peut accéder aux informations soit en sélectionnant une commune soit en entrant le numéro d'un programme. La carte d'implantation des programmes de Bouygues Immobilier apparaît automatiquement sur la droite. Une fois la commune ou le programme sélectionné, un système de « cross-filter » permet de cumuler des critères - chiffres clés (données INSEE), indicateurs thématiques et informations sur les programmes – mettant à jour automatiquement les indicateurs positionnés sur la carte.

Pour Solène Touret, Responsable stratégie et innovation commerciale et pilote du projet MyGeoMarket chez Bouygues Immobilier, « cet outil nous aide à l'élaboration de notre communication commerciale et représente un gain de temps considérable en valorisant des données que nous sous-exploitions ». Lorsqu'auparavant le marketing n'avait accès qu'à de la connaissance partielle, avec des données réparties dans diverses bases, MyGeoMarket réunit en une seule plateforme données commerciales, géographiques et clientèle, et permet de baser les prises de décision sur des éléments factuels et à jour.

« MyGeoMarket permet une double approche intéressante pour nos métiers. Nous pouvons mieux appréhender notre cible dans ses habitudes de vie mais aussi effectuer une meilleure analyse des réservataires », souligne Alice Bartaux, Responsable Marketing.

La première version de MyGeoMarket a été mise à disposition des collaborateurs de Bouygues Immobilier en mai 2016. Une seconde version est d'ores et déjà envisagée, avec l'ajout d'autres sources de données. MyGeoMarket marque la rencontre de trois expertises, mises en synergie, pour créer un outil performant, et le début d'un partenariat fort entre Synaltic, Publicis Media et Bouygues Immobilier.





## COMMENT VOYEZ-VOUS L'ÉVOLUTION DU BIG DATA ? VA-T-ON VERS PLUS DE SMART QUE DE BIG DATA ?

Pour moi, le Big Data évolue actuellement dans trois directions :

### La complexité de l'infrastructure

Le Big Data réclame un certain investissement en nouvelles connaissances et compréhensions. Si des éléments restent encore flous, les nouveaux prérequis tels que la scalabilité, l'élasticité, la tolérance à la panne, la haute disponibilité et bien sûr la sécurité semblent être de plus en plus maîtrisés, surtout par les DSI. Et ce n'est qu'une automatisation d'architecture qu'elles connaissaient déjà. La complexité qui en découle profite clairement aux offres directement érigées dans le cloud. Seules les entreprises dont le métier réclame une maîtrise complète de l'infrastructure et du réseau résisteront au cloud. Le dernier Hadoop Summit à San Francisco, que nous avons relaté sur notre blog, montrait à quel point cela devenait un des plus grands enjeux de notre secteur.

### Le morcellement de l'offre

Les problématiques de gestion de données sont très larges et l'on voit ainsi des solutions de plus en plus focalisées sur des sujets très précis. C'est, d'ailleurs, un des soucis de l'écosystème Hadoop par exemple : les entreprises utilisatrices se sentent perdues et démunies face à une offre de plus en plus morcelée. Synaltic se plonge régulièrement dans ces méandres pour fournir les architectures les mieux adaptées aux besoins de ces clients.

### Le Self-Service

Nous avons rédigé un livre blanc sur la Data Preparation, qui devient une discipline à part entière au sein de la chaîne de valeur de la BI. Tableau avait déjà bouleversé ce marché avec la Data-Visualisation. La Data Préparation, couplée au référentiel des jeux de données qui s'inscrivent dans la gouvernance collaborative de la donnée, sont doucement en train de transformer le marché de l'informatique décisionnelle. Et ce même en Big Data.

## QUELLE EST LA MATURITÉ DES ENTREPRISES FRANÇAISES ? Y A-T-IL DES SECTEURS PLUS AVANCÉS ?

Nous rencontrons de nombreuses entreprises : des grandes, des petites... De la volumétrie, des données de toutes sortes, des données en flux continu, elles en ont. Mais leurs architectures ne semblent plus suffire. En effet, nombreuses sont les entreprises qui font appel à nous pour les guider dans ce domaine : elles ont déjà mis en place elles-mêmes leurs architectures dédiées au Big Data mais rencontrent malheureusement des difficultés.

Globalement, Hadoop et les technologies NoSQL ont une dizaine d'années aujourd'hui. Plus aucun pan de l'informatique ignore ces technologies, pas même le plus petit acteur ! Et nos clients non plus ! Ils savent au moins ce que c'est. Et s'ils ne les ont pas encore intégrées, ces technologies font partie de leur feuille de route, à moyen ou à long terme.

Il existe diverses manières d'adopter ces technologies, via un data lake, dans une démarche de digitalisation ou plutôt de mise en données de son organisation (datafication), dans la construction d'applications plus réactives, dans une démarche de self-service en informatique décisionnelle couplée à une réelle démarche de gouvernance des données. Où qu'elles en soient dans leur démarche, quels que soient leurs objectifs, nous accompagnons les entreprises de toutes tailles pour effectuer les choix les plus éclairés avec les technologies les plus adaptées.

## VOTRE CELLULE R&D EST EN VEILLE SUR LES DERNIÈRES TENDANCES TECHNOLOGIQUES, POUVEZ-VOUS NOUS DÉVOILER QUELQUES EXCLUSIVITÉS ?

Il y a une véritable tendance de fond, orientée vers la création de plateformes qui gèreraient la donnée de bout en bout. D'autant plus si l'on se place au niveau de l'agrégation de données en temps réel. Il y a une grande volonté de réduire la complexité des modes de traitement de la donnée avec une vision plus réactive, tant pour fournir des indicateurs, que dans la maintenance elle-même (développement et capacité à les faire évoluer rapidement, pour les livrer tout aussi rapidement).

On ne parle plus que de logs, désormais. Mais attention, il ne s'agit pas en fait de traces d'erreurs, mais simplement d'une information granulaire, unitaire. Celle-ci, récoltée régulièrement, s'agglomère de manière continue pour délivrer les chiffres les plus à jour.

Pour aider nos clients dans cette logique, nous avons mis au point une solution pour mieux gérer les échanges de données et les monitorer. Bien sûr nous avons veillé à prendre en compte les outils de Big Data pour répondre à ces contraintes.

**Edouard BEAUCOURT**

Directeur Régional France et Europe du Sud

+33(0)1 70 91 88 94

salesparis@tableau.com

La Madeleine - 19 boulevard Malesherbes 75008 Paris



[www.tableau.com/fr](http://www.tableau.com/fr)



## CHEZ UNILEVER, LES DONNÉES, VISUALISÉES GRÂCE À TABLEAU, DÉVOILENT LES COMPORTEMENTS DES CONSOMMATEURS

À l'échelle internationale, Unilever est une entreprise extraordinairement complexe et l'architecture l'est également. Le groupe distribue des produits dans différentes catégories dans plus de 170 pays et fait face à des concurrents de natures diverses. Certains des concurrents commercialisent des produits spécialisés ciblant un seul marché, d'autres sont des multinationales comme Unilever qui vendent de très bons produits dans le monde entier. De plus en plus d'individus s'intéressent aux données, pas uniquement pour réduire la complexité de l'entreprise à l'échelle mondiale en fournissant simplement une vue agrégée, mais ils cherchent également à comprendre les tendances de consommation avec un niveau de détail bien plus important.

Dans le cas d'Unilever, de nombreuses entreprises avaient leur propre infrastructure informatique, et quand ces entreprises ont rejoint le groupe suite à des acquisitions ou d'autres programmes de synergie, il a fallu trouver un moyen d'intégrer leurs technologies locales à la structure globale du groupe. En effet, le directeur du service analytique a dû faire face à un nombre de systèmes incalculable. Du point de vue des données, de l'acquisition des données et de l'infrastructure informatique, la situation est plutôt complexe.

L'équipe Walmart chez Unilever aux États-Unis est très axée sur les données. Elle sait poser les bonnes questions et utilise les données chaque jour pour répondre à ces questions. Les membres de l'équipe ont demandé au service IT s'il pouvait trouver un outil pour les aider à accéder aux données plus facilement et permettre à ceux qui se posent des questions d'obtenir des réponses. Après avoir essayé Tableau pendant 10 minutes, le service IT a compris que cet outil allait changer les règles du jeu. Au terme d'une période d'essai probante, Tableau fut peu à peu lancé dans toutes les unités opérationnelles aux États-Unis. La solution fut alors présentée à Rishi Kumar, directeur du service analytique en sachant qu'il adorait découvrir de nouveaux outils. « Mon collègue m'avait dit : « Si tu es accro à Excel, tu vas vite devenir un fan de Tableau ». Il avait raison. »

Aujourd'hui, Tableau permet au groupe d'étudier ses données aussi bien au niveau international que local. Les responsables internationaux l'utilisent pour explorer les données au niveau de détail souhaité pour comprendre l'activité et prendre les décisions nécessaires.

Tableau est devenu indispensable pour les aider à bien comprendre ce qui pousse les consommateurs à effectuer certains achats et comment ces facteurs évoluent dans le temps, en fusionnant toutes les données.

« Tableau est excellent pour explorer les données en détail, et il permet également d'obtenir des perspectives différentes, ». Les utilisateurs peuvent ainsi effectuer des regroupements sur des produits ou des concurrents, selon les cas. Et cette possibilité de regrouper les données à ce niveau de détail est un atout majeur pour la société. « Tableau est aussi très évolutif ; nous n'avons donc pas besoin d'investir dans différentes infrastructures informatiques à l'échelle mondiale : une seule suffit désormais. »

Avec l'utilisation de Tableau beaucoup de choses ont changé. En effet, de nombreuses opérations qui nécessitaient des formules mathématiques complexes sont désormais un jeu d'enfant grâce à la fonction de visualisation des données de Tableau. Plutôt que de s'appuyer sur des ordinateurs et des modèles mathématiques complexes pour traiter les informations et décrire leurs relations, Tableau a permis de visualiser ces données et de laisser l'esprit humain les interpréter et comprendre leurs relations. « Jusque-là, il nous fallait des experts qui maîtrisaient le côté mathématique des statistiques et de l'analyse, mais désormais des utilisateurs lambda dans l'entreprise sont capables de visualiser simplement les données et d'affûter leur esprit d'analyse. »

De plus, poser des questions ravive l'intérêt des gens. Dans les grandes entreprises en particulier où le rythme est effréné, la tendance est de simplement suivre le processus. Se poser des questions sur le pourquoi de la situation, sur ce qui se passe vraiment est rare. Seuls, les chiffres comptent. Une suite de causes justifiant la situation est enregistrée et un rapport et transmis jusqu'à ce qu'il donne lieu à une prise de décision.

« Avec Tableau, chacun a la possibilité de poser des questions et d'obtenir des réponses, ce qui explique ce regain d'intérêt. On constate que les utilisateurs sont plus curieux vis-à-vis de l'analyse. »

Le second changement observé est la capacité d'illustrer des situations compliquées d'une manière plus simple, en supprimant toute complexité. « Avec Tableau, on peut par exemple facilement regrouper tous les magasins fréquentés aux États-Unis en majorité par des clients d'origine hispanique ou regrouper tous les produits qui enregistrent une forte croissance. »

Au quotidien, Tableau permet à Unilever d'utiliser les tendances locales, à petite échelle, basées sur le comportement du consommateur pour définir une espèce de tendance globale, de dynamique du marché.

« Par exemple, les déodorants à bille se vendent mieux que les déodorants en spray. Nous cherchons à savoir pourquoi. Je pense qu'être capable de comprendre la dynamique du marché en tenant compte de ses éléments constitutifs, aussi petits soient-ils, en particulier les biens de consommation, représente un avantage concurrentiel, car en fin de compte, ce sont les consommateurs, leurs comportements et les tendances de la consommation qui influencent véritablement les performances à un niveau agrégé. »

Enfin, Tableau a permis à Unilever de raconter des histoires, notamment du fait d'un contenu attrayant. Les utilisateurs peuvent voir des choses qu'ils n'avaient jamais vues auparavant. Ils peuvent vérifier les prix moyens dans le monde ou identifier les concurrents ayant fait leur apparition sur certains marchés.

Tableau peut regrouper un nombre incalculable d'informations dans un seul visuel. Réunies dans un PDF, ces mêmes informations seraient difficilement exploitables. « Je suis convaincu que nous pourrions visualiser toujours plus de données et susciter l'intérêt de davantage d'utilisateurs, car ces présentations visuelles riches d'informations nous permettent de raconter des histoires. »



# Edouard Beaucourt

Directeur Régional France et Europe du Sud

## INTERVIEW



Bien comprendre l'usage des données dans le domaine professionnel, avec Edouard Beaucourt

### COMMENT L'ÉVOLUTION DU MARCHÉ DE LA DATA A ÉVOLUÉ CES DERNIÈRES ANNÉES ?

Les entreprises ont aujourd'hui accès à un volume de données sans précédent. Chaque jour, un volume de près de 2,5 milliards de Go est créé. IDC estime qu'en 2020 le monde générera 50 fois plus de données qu'en 2011 et comptera 75 fois plus de sources d'information. Pour une grande entreprise, le volume de Big Data se chiffre en Po ou au-delà, alors que pour une entreprise de taille moyenne, un volume se chiffrant en dizaines ou en centaines de To peut être complexe à analyser et à gérer.

Les données constituent pourtant un outil très puissant, à condition de pouvoir en extraire des informations exploitables. Les organisations sont submergées de données provenant de différentes sources, qu'il est nécessaire d'interpréter rapidement pour prendre des décisions en conséquence. Alors qu'il est primordial de prendre des décisions importantes très rapidement en raison de l'évolution rapide des marchés, il devient nécessaire de tirer parti de cette abondance des données afin de pouvoir rapidement les analyser et identifier des informations exploitables qui deviendront une source de valeur pour l'entreprise.

### COMMENT TABLEAU AIDE JUSTEMENT À RÉPONDRE À CETTE ÉVOLUTION ?

Nous aidons les entreprises à voir et à comprendre leurs données. Tout ce que nous faisons repose sur ces quelques mots. Nos produits changent la façon dont les utilisateurs utilisent les données pour résoudre leurs problèmes.

Tableau permet, en fait, aux entreprises grâce à l'utilisation de représentations graphiques attrayantes et interactives de mieux visualiser et comprendre les données. L'analyse visuelle consiste à transformer les données en une représentation graphique, comme un diagramme, une courbe ou une carte, pour permettre de comprendre les tendances instantanément. La représentation graphique se met à jour lorsqu'une variable est modifiée, ce qui facilite la présentation de l'histoire des données en temps réel. Nous rendons l'analyse des données rapide et simple, esthétique et utile. Nos logiciels sont destinés à tout le monde.

### DE QUELLE MANIÈRE VOS OUTILS PERMETTENT-ILS DE SOUTENIR LE TRAVAIL DES ENTREPRISES ?

Le fait est qu'il est plus facile de comprendre des données lorsqu'elles sont présentées visuellement. Les utilisateurs réfléchissent de manière visuelle et peuvent rapidement prendre des décisions en fonction de leurs données lorsqu'ils sont en mesure de les visualiser et de les comprendre. Par ailleurs, un outil adapté permet à tout utilisateur dans l'entreprise d'effectuer des analyses, sans pour autant qu'il soit nécessaire d'être un pro des statistiques ou de maîtriser la science des données.

Un tel degré d'accessibilité est d'autant plus crucial que le volume actuel de données impose à chaque collaborateur, et donc pas uniquement aux analystes à part entière, de se familiariser avec l'analyse des données, alors que les produits d'analyse traditionnels ont été conçus pour centraliser et contrôler les données, et non pour démocratiser leur utilisation. La majorité des entreprises se retrouve alors à devoir exploiter des feuilles de calcul labyrinthiques pour contourner ces systèmes rigides, ou doit faire appel à des spécialistes des données pour trouver des réponses à des questions basiques.

Une telle approche ne présente pas d'avantages. Il est plus avantageux de donner à chacun la capacité de réfléchir sur les données et de prendre des décisions en conséquence. L'analyse de données en libre-service décuple les possibilités des employés et leur permet de déceler les informations exploitables dont ils ont besoin pour améliorer leurs performances et leur productivité. Ainsi, les collaborateurs participent au développement de l'entreprise d'une manière entièrement innovante.

### COMMENT CELA SE CONCRÉTISE-T-IL EN PRATIQUE ?

Les organisations, en mesure de tirer au mieux parti des données dans le processus décisionnel, ont tendance à connaître une meilleure réussite. Le déploiement d'une solution, comme celles de Tableau, permet de déceler les tendances et les données incohérentes, d'identifier les manières de réduire les coûts, de renforcer l'efficacité et de déceler les domaines à améliorer. Les entreprises à la pointe de l'exploitation de données sont trois fois plus susceptibles d'avoir également un avantage financier important sur la concurrence. Par conséquent, permettre aux utilisateurs de visualiser les données n'est pas uniquement utile pour la forme, c'est également indispensable pour permettre à l'entreprise de rester compétitive sur le marché actuel. Sans une telle capacité, une entreprise disposera simplement d'entrepôts de données sans les outils pour les exploiter, et accusera par conséquent un retard certain sur ses concurrents.

Testez gratuitement la dernière version de Tableau en vous connectant sur notre site : [www.tableau.com/fr](http://www.tableau.com/fr)

**Marthe BUISSON**  
 FIELD MARKETING MANAGER (FRANCE)

+33(0)1 46 25 06 00  
 mbuisson@talend.com  
 9 - rue Pages 92150 Suresnes France



www.talend.com



**CHIFFRES CLEFS**

690 milliards d'euros de transactions inter-entreprises  
 230 millions d'expériences de paiement  
 120 000 utilisateurs dans 65 pays  
 Gain de temps = un facteur de 3 à 4 avec Talend

**LE BIG DATA POUR ACCÉLÉRER LA GÉNÉRATION DE CASH-FLOW DES ENTREPRISES QUAND LE BIG DATA SE MET AU SERVICE DES DIRECTIONS FINANCIÈRES**

Sidetrade utilise la technologie 'Talend Real-Time Big Data' afin de proposer à ses clients une nouvelle plateforme d'analyse prédictive des paiements inter-entreprises.

**SE FAIRE PAYER EN RETARD PAR SES CLIENTS EST-IL UNE FATALITÉ ?**

Selon les années et les pays, la plupart des entreprises subissent un retard de paiement d'environ 13 à 15 jours. Sidetrade est un éditeur de logiciels dans le Cloud, spécialisé dans l'amélioration de la performance du cycle Sales-to-Cash, c'est-à-dire de la vente jusqu'au paiement client. Nous offrons une vision à 360° de la relation client permettant un contrôle et une analyse prédictive des flux financiers clients. Sur les trois dernières années, nous avons enregistré plus de 690 milliards d'euros de transactions inter-entreprises, ce qui représente à peu près 230 millions d'expériences de paiement. L'objectif, pour Sidetrade était d'utiliser les millions de données que nous avons dans notre Cloud.





### ET C'EST LA RAISON POUR LAQUELLE VOUS AVEZ MIS À DISPOSITION DE VOS CLIENT UNE PLATEFORME D'ANALYSE PRÉDICTIVE ?

Sidetrade Payment Intelligence (SPi) est un score de paiement prédictif qui permet de prévoir le comportement de paiement des clients. Nous avons créé une Application, dénommée SPi Solution, où les directions financières peuvent confronter leur performance en matière de délai de paiement au SPi correspondant et ce à tous les niveaux de l'entreprise : à l'échelle d'un Groupe, d'un Pays, d'une Business Unit ou d'un client individuel en particulier. Elles peuvent ainsi définir des objectifs de génération de cash à 3, 6, 9 ou 12 mois et être en mesure de les suivre et faire un suivi dans le temps. Elles peuvent agir soit sur une diminution des délais de paiement accordés en collaboration avec les équipes commerciales, soit en comparant la performance des équipes Finance sur la maîtrise des seuls retards de paiement par rapport à ceux subis par les autres fournisseurs sur ces mêmes clients. La performance dans la négociation des délais accordés ou dans la maîtrise des retards subis devient donc objective et éclaire les Directions Financières sur leur potentiel d'amélioration dans la génération future de cash.

### QUEL RÔLE JOUE LE BIG DATA DANS CETTE PLATE-FORME ?

Le Big Data est important pour une plate-forme d'analyse prédictive comme celle de Sidetrade puisque les volumes de données que nous manipulons pour établir les scores prédictifs de comportement de paiement sont absolument gigantesques. Nous nous appuyons sur une architecture massivement parallèle reposant sur Hadoop afin de traiter, de manière anonyme, les téraoctets représentés par les millions d'expériences de paiement enregistrées dans notre Cloud.

### COMMENT TALEND BIG DATA VOUS A-T-IL AIDÉ À METTRE EN PLACE CETTE PLATE-FORME ?

Talend Big Data nous apporte des compétences Hadoop rares. En simplifiant l'appropriation d'Hadoop, la solution de Talend répond parfaitement à nos exigences. Elle génère nativement du code Hadoop sans passer par des composants propriétaires ; la solution est compatible avec toutes nos bases de données relationnelles et se distingue par une prise en main rapide. Talend s'intègre très simplement dans l'architecture Sidetrade existante : nous avons des connecteurs vers Oracle, vers MongoDB, vers SQL Server, et bien entendu vers Hadoop, tout ceci de manière totalement transparente et par drag-and-drop... On estime que 40 à 60 % du temps total d'un projet Big Data est passé à l'extraction, au nettoyage et à l'unification de ces données sans parler de la génération du code natif dans Hadoop. Grâce à Talend, nous avons pu réduire ce délai par un facteur entre 3 et 4. La solution Talend Big Data nous permet de travailler dans une interface user-friendly tout en générant du code natif dans Hadoop - ce qui représente pour nous un gain de temps significatif puisque tous les mois, nous collectons des millions de données à enregistrer. Au-delà du fait de démocratiser l'accès aux technologies Hadoop, la réduction des coûts de stockage et le traitement extrêmement rapides des données dans le cloud de Sidetrade, Talend s'assure également de la qualité des données utilisées pour l'analyse.

### QUELS SONT LES BÉNÉFICES POUR LES DIRECTIONS FINANCIÈRES ?

Avec cette nouvelle plateforme, Sidetrade fournit une solution tout-en-un d'anticipation des comportements de paiement afin de simplifier la vision des dirigeants d'entreprises et responsables financiers, et de leurs permettre de disposer d'une information prédictive unique sur la quasi-totalité de leurs clients. Cette innovation contribue à améliorer l'efficacité de la finance et du commerce en fournissant aux départements financiers un contrôle temps réel sur le business ainsi que des opportunités de revenus. Les directions financières peuvent comparer leurs performances en matière de délais de paiement - accordés ou subis - dans une approche fondée sur l'analyse de données et non sur une intuition. La mise à disposition d'une Plate-Forme d'Analyse Prédictive sur les comportements de paiement à base de machine learning est une première dans l'industrie du crédit interentreprises. Du point de vue de la génération de Cash, l'enjeu est significatif : en fonction de votre chiffre d'affaires, un simple jour de délai de paiement gagné peut représenter plusieurs millions d'euros de cash additionnel. À partir du moment où vous avez une connaissance très fine et prédictive de la façon dont votre client vous règle ses factures, l'effet est double : d'une part, vous agissez pour raccourcir vos délais et générez du cash, d'autre part, vous améliorez votre satisfaction client ! C'est pour ça que le cycle Sales-to-Cash doit être envisagé de bout en bout : depuis la négociation des termes de paiement jusqu'à l'encaissement des fonds.



**+33(0)1 45 81 81 00**  
 masteres@telecom-paristech.fr  
 46 rue Barrault 75634 Paris Cedex 13 France



[www.telecom-paristech.fr/bigdata](http://www.telecom-paristech.fr/bigdata)



## LE BIG DATA, UNE EXPERTISE AU CŒUR DE TELECOM PARISTECH

Première grande école française d'ingénieurs dans le domaine des Technologies et des Usages du Numérique, Télécom ParisTech forme ses diplômés à innover et entreprendre dans un monde désormais numérique. Près de 1500 étudiants la choisissent chaque année dont 55 % d'internationaux. Toutes les disciplines fondamentales des Technologies et Usages du Numérique (y compris les Sciences Economiques et Sociales), sont couvertes par ses enseignements et sa recherche.

Télécom ParisTech a une forte prise sur le monde socio-économique et le pourvoit en ingénieurs, docteurs et professionnels, formés pour accompagner la transformation numérique des entreprises, de l'industrie et de la société. Télécom ParisTech transfère à l'industrie les résultats de sa recherche de très haut niveau, via des brevets, des licences et des start-up, notamment au sein de son incubateur, ParisTech Entrepreneurs.

École de l'Institut Mines-Télécom, membre fondateur de ParisTech et de l'Université Paris-Saclay, Télécom ParisTech se définit comme le Collège de l'innovation par le numérique de Paris-Saclay, dont l'ambition est de devenir l'un des premiers pôles d'innovation mondiaux.

## UNE RECHERCHE PLURIDISCIPLINAIRE UNIQUE

Portée par 170 enseignants-chercheurs permanents et 300 doctorants, sa recherche, de réputation internationale est évaluée A+ par l'AERES et est couronnée chaque année par des prix prestigieux (Médaille d'Argent CNRS, ERC, Google Awards, etc.). Ses laboratoires allient des expertises disciplinaires de très haut niveau dans les sciences et technologies de l'information et une capacité rare de mobilisation transversale interdisciplinaire.

Le dynamisme de la filière Big Data à Télécom ParisTech découle d'une recherche pluridisciplinaire autour des mathématiques appliquées, de l'informatique et des sciences humaines et sociales. Elle réunit une cinquantaine d'enseignants-chercheurs et autant de doctorants. En partenariat avec une quinzaine d'entreprises, l'école s'investit dans trois chaires de recherche dans le domaine du Big Data.

La chaire Machine Learning for Big Data conduit ses recherches à l'interface des mathématiques et de l'informatique. La machine learning vise à élaborer des algorithmes permettant aux machines d'apprendre automatiquement à partir des données et ainsi d'améliorer leurs performances. La chaire Big Data & Market Insights regroupe des chercheurs spécialisés dans la gestion et la fouille de données massives, l'extraction de connaissances à partir du Web, l'analyse de réseaux sociaux et le marketing digital.

La chaire Valeurs et Politiques des Informations Personnelles traite des aspects juridiques, techniques, économiques et philosophiques qui concernent la collecte, l'utilisation et le partage des informations personnelles.

## UN ENVIRONNEMENT D'ENSEIGNEMENT EXCEPTIONNEL

Télécom ParisTech a bâti un écosystème unique, basé sur une forte interaction entre formation, recherche et innovation et qui se caractérise par l'accueil régulier de conférences et de colloques internationaux, un incubateur qui héberge plusieurs start-up issues du Big Data, des partenariats avec les entreprises clé du secteur...

Cet environnement a permis au Mastère Spécialisé® « Big Data » d'être le premier du genre en France et de faire figure de référence dans le domaine. Par la suite, l'école a mis en place une large gamme de formations pour tous les besoins, notamment dans le cadre de Télécom Evolution. Télécom ParisTech entretient des liens étroits avec l'industrie : une dizaine d'entreprises sont membres du comité de veille et de perfectionnement des formations et plusieurs dizaines participent aux enseignements, études de cas, mises en situation professionnelle, tables rondes et séminaires. Elles proposent également stages, thèses professionnelles, projets de groupe mais aussi données réelles anonymisées.

## PRIORITE A L'INNOVATION

L'incubateur ParisTech Entrepreneurs a accueilli en 15 ans plus de 300 start-up innovantes du numérique. Beaucoup d'entre elles utilisent les technologies du Big Data, que ce soit dans l'analyse et le traitement des données, le marketing et le e-commerce, l'industrie, les transports, la santé, la high-tech, les services Internet, la finance... L'entrepreneuriat est en effet fortement encouragé au sein des différentes filières de l'école. Télécom ParisTech participe également à la plateforme souveraine Teralab qui a pour but de lever la barrière de l'adoption des technologies Big Data. Infrastructure puissante, flexible et de confiance, Teralab accompagne ses utilisateurs sur les plans technologie et juridique. Elle héberge des projets portés par les entreprises, les start-up ou les organismes de recherche et de formation.

## TELECOM EVOLUTION, DES COMPETENCES TOUJOURS A LA POINTE

Télécom Evolution est le centre de formation continue commun à Télécom ParisTech, Télécom Bretagne et Télécom SudParis. Il accompagne les entreprises et les organisations dans la montée en compétences de leurs collaborateurs autour des grandes problématiques du numérique : Internet des objets, Big Data, cybersécurité, transports intelligents... Télécom Evolution propose une gamme complète de formations depuis l'initiation jusqu'à l'expertise, dans une grande variété de modalités pédagogiques, du MOOC à la formation certifiante, dispensés par les meilleurs experts et des professionnels reconnus.



Les tables rondes entreprises font salle comble



Jack Dorsey, créateur et directeur général de Twitter, répondant aux questions des étudiants

## Mastère Spécialisé® Big Data : Gestion et analyse des données massives

Un Bac+6 pour des diplômés en poursuite d'études ou en reconversion, accrédité par la Conférence des Grandes Écoles. **Plus de 600 heures sur 9 mois de cours**, travaux pratiques, séminaires, et un projet de groupe « *fil rouge* » proposé par une entreprise.

Un stage et une thèse professionnelle en entreprise sur 4 à 6 mois concluent un programme **conçu et enseigné par des spécialistes du domaine**, en forte adéquation avec les besoins des entreprises.

Le MS Big Data combine un large socle de connaissances techniques et une approche complète des aspects juridiques, économiques, business.

**100 % des diplômés** ont trouvé un travail facilement, dont **89% en moins de 2 mois**.\*

\*Étude Harris Interactive, novembre 2015

TELECOM  
ParisTech



MS  
Master spécialisé®



## CES Data Scientist (Certificat d'études spécialisées)

**Pour les professionnels en poste** qui souhaitent une certification en science des données (collecte, stockage et calcul distribués, analyse statistique, visualisation).

**Très opérationnel**, le CES permet la maîtrise des techniques de gestion et d'analyse des Big Data et des principaux algorithmes du machine learning.

**12 sessions de 2 jours sur 10 mois** : cours, travaux dirigés, travaux pratiques, témoignages de professionnels et une plateforme d'e-learning.

## MOOC Fondamentaux pour le Big Data

**Gratuit, flexible et indispensable** pour valider une inscription au **Mastère Spécialisé®** ou au **CES**.

**En 6 semaines**, révisez les prérequis essentiels dans les domaines de l'analyse, de l'algèbre, des probabilités, des statistiques, de la programmation Python et des bases de données.

**7 modules**, des vidéos, des quizz et un projet final, et plus de **7000 inscrits par session** !

TELECOM  
Evolution



## Stages courts (2 jours)

- Big Data : premiers succès et retours d'expérience
- Big Data : enjeux stratégiques et défis technologiques
- Big Data : panorama des infrastructures et architectures distribuées
- Data Science dans le Cloud : Big Data, statistiques et Machine Learning
- Data Science : introduction au Machine Learning
- Introduction à la sécurité du Big Data
- Visualisation d'information (InfoVis)
- Moteurs de recommandation et extraction de données du Web
- Text-Mining
- Introduction à R



## Executive MBA Data Scientist des Métiers de l'assurance

Conçu **en partenariat avec l'École Polytechnique d'Assurances**, cet Executive MBA s'adresse à des ingénieurs, techniciens, chefs de projet, informaticiens, statisticiens, mathématiciens souhaitant développer leurs compétences dans le domaine du Big Data appliqué à l'assurance.

La formation compte **67 jours en présentiel sur 13 mois** à raison d'une semaine par mois. Le programme repose sur **4 piliers** : technique (Data Science et Assurance), stratégique, éthique et management. La validation de la partie Data Sciences du MBA donne lieu à la **délivrance par Télécom ParisTech du Certificat d'Études Spécialisées (CES) Data Scientist**.



EP  
ÉCOLE POLYTECHNIQUE D'ASSURANCES

Toutes nos formations : [www.telecom-paristech.fr/bigdata](http://www.telecom-paristech.fr/bigdata)



Teradata aide les entreprises à générer plus de valeur à partir des données quel que soit leur format et provenance. Nos solutions analytiques multi-genre, les architectures Big Data flexibles que nous proposons et nos experts aideront votre société à obtenir un avantage compétitif durable grâce à l'analyse des données.

Vous utilisez ainsi toute la valeur et la pertinence tirées de l'analyse des données pour être plus performant, plus agile, plus innovant et encore mieux servir vos clients et votre stratégie.

Teradata Solutions

Technology Solutions

Public Cloud, Private Cloud, Managed Cloud, Hybrid Cloud, On-premises  
 Teradata Database, Aster Analytics, Hadoop  
 Teradata QueryGrid, Presto, Listener, Unity, AppCenter



Business Solutions

Analytic Solutions  
 Analytics Business Consulting  
 Business Value Framework  
 Data Science

Architecture Expertise

Strategy and Roadmaps  
 Design and Implementation  
 Ecosystem Architecture  
 Managed Services

High-impact Business Outcomes

SIEMENS : RENDRE LES SERVICES FERROVIAIRES FIAbles EN COMBINANT ANALYTIQUE ET BIG DATA

L'importance du Big Data

La division Siemens Mobility Data Services tire parti du Big Data et des solutions analytiques pour s'assurer que les services de transport qu'elle livre dans le monde entier soient rapides, fiables et plus économes en énergie. Siemens produit une gamme de trains et de composants d'infrastructure, tels que des systèmes électriques, d'automatisation, de signalisation et de contrôle ferroviaire. La division MDS propose des solutions pour la mobilité urbaine/interurbaine et la logistique (le trafic ferroviaire longue distance est un autre secteur important de l'entreprise).

Big Data : une architecture évolutive

Grâce à la présence de dizaines de milliers de capteurs, Siemens peut désormais utiliser les données qu'ils génèrent en association avec celles provenant de l'ensemble de l'entreprise et de sources des données externes, comme la météo, pour relever leurs défis métiers de manière novatrice. Mais avec l'important volume de données non structurées, de données de capteur et de données télématiques, Siemens avait besoin d'une architecture évolutive performante. Les données provenant des trains et des rails, des processus de réparation, les données météorologiques et de la chaîne d'approvisionnement se retrouvent toutes dans Hadoop, Teradata Aster et Teradata Data Warehouse qui sont les composants de Teradata Unified Data Architecture.

« Nous ne pourrions pas faire ce que nous faisons avec une architecture différente car les volumes de données que nous traitons sont très importants », a souligné Gerhard Kress, Directeur, Siemens Mobility Services. Quel volume cela représente-t-il ? « Lorsque nous rassemblons toutes les données des capteurs d'une seule flotte de trains en Europe, cela gé-

nère 100 milliards de lignes d'une table », cite Gerhard Kress. « Si vous voulez exécuter un algorithme de Machine Learning sur un tel volume, cela ne fonctionne que si il est exécuté de manière massivement parallèle, d'où l'importance de cette architecture ».

Des solutions analytiques Big Data de pointe

Chez Siemens, les data scientists et les ingénieurs examinent les données provenant de l'ensemble de l'entreprise, et les analysent à l'aide des dernières techniques et outils d'analyse, comme le Machine Learning, pour anticiper de manière précise les défaillances réelles de pièces, tout en identifiant aussi les faux positifs (la prévision d'une défaillance qui ne se produit pas).

« Nous disposons de données et de modèles d'analyse nous permettant de réellement prévoir certaines défaillances. Par exemple, les boîtes de vitesses des trains à grande vitesse sont parmi les éléments les plus délicats à surveiller », a déclaré Gerhard Kress. « A plusieurs occasions, nous avons pu prévoir que ces systèmes seraient défaillants au cours des semaines suivantes. Nous avons alors eu le temps pour fournir les pièces de rechange, les réparer, retirer le train de la circulation sans nuire à la planification et travailler avec le client, sans le moindre problème ».

Voici un autre exemple qui démontre le niveau de sophistication des analyses faites par Siemens. En analysant des données météorologiques, ils peuvent différencier ce qui semble plus susceptible de tomber en panne sur un train à grande vitesse circulant entre Moscou et Saint-Pétersbourg pendant un hiver glacial par rapport à un train à grande vitesse circulant en Espagne pendant un été torride.

L'avantage du Big Data

Grâce à la combinaison d'une architecture Big Data appropriée et d'une analytique sophistiquée, Siemens peut désormais :

- Prédire avec exactitude quand des moteurs sont susceptibles de tomber en panne
- Assurer une chaîne d'approvisionnement de pièces efficace afin de réduire ou d'éliminer les temps d'arrêt
- Évaluer la façon dont des conditions météorologiques sont susceptibles d'affecter le même modèle de train dans différentes régions

Grâce au Big Data, Siemens peut maintenant garantir des horaires de trains extrêmement fiables. Pour un trajet Barcelone-Madrid, 80% des voyageurs avaient l'habitude de prendre l'avion, contre seulement 30 % aujourd'hui. Pas plus d'un voyage sur 2 000 en train à grande vitesse connaît un retard de cinq minutes maximum. Siemens assure aussi les services du métro de Bangkok où seulement 1 % des trajets connaissent un retard.



Vous voulez en savoir plus à propos de l'architecture Big Data et des outils d'analyse utilisés proposés par Teradata ? Les experts Teradata seront ravis de répondre à vos questions sur notre stand 421.

Retrouvez tous nos retours d'expérience clients sur notre site [www.teradata.com](http://www.teradata.com)



# La Prochaine Grande Évolution se trouve Déjà dans vos Données

Chaque domaine de votre business génère des données. Le monde qui vous entoure génère des données. Chaque donnée vous donne une vision partielle de votre business. Vous n'aurez donc une vision complète que lorsque vous analyserez l'ensemble des données internes et externes. Extraire les données des silos applicatifs dans lesquels elles sont stockées, les analyser ensemble avec les données externes, vous aidera à mieux appréhender votre business. Des sociétés telles que Siemens, Volvo et JD Williams, analysent l'ensemble de ces données pour prendre des décisions efficaces. La prochaine grande évolution de votre entreprise se trouve déjà dans les données. Découvrez la dès aujourd'hui!



TIBCO Software accompagne les entreprises dans leur transformation digitale en interconnectant, en temps réel, les différents éléments de son système d'information et en délivrant davantage d'intelligence aux collaborateurs de l'entreprise, de l'utilisateur métier au data scientist. Cette combinaison garantit des réponses rapides, de meilleures décisions et des actions pertinentes. Depuis près de 20 ans, des milliers d'entreprises font confiance à la technologie de TIBCO pour se différencier en offrant à leurs clients des expériences convaincantes, des actifs optimisés et des business modèles innovants.

## INTERCONNECT EVERYTHING



### TOUT INTERCONNECTER

La plateforme digitale TIBCO offre les fonctionnalités requises pour être le socle reliant systèmes, personnes et objets :

- **Intégration** : les solutions d'intégration et de messagerie connectent les systèmes de l'entreprise et de son écosystème pour créer des services fournissant le niveau adéquat de visibilité ou d'action.
- **Gestion des processus métier – BPM & Case Management** : les solutions de gestion des processus métier permettent d'impliquer l'ensemble des acteurs liés aux activités de l'entreprise - clients, partenaires ou collaborateurs - dans ses processus clés.
- **Gestion des API** : les API s'imposent comme un support des stratégies marketing et commerciale. La solution TIBCO Mashery® Entreprise permet l'accès aux données depuis les smartphones, navigateurs et autres terminaux. La mise à disposition de services numériques est garantie à l'ensemble de l'écosystème de l'entreprise sous une forme ouverte, pilotée et sécurisée.

### INTELLIGENCE AUGMENTÉE

La plateforme digitale TIBCO permet à tous les collaborateurs de l'entreprise, de disposer d'une vision pertinente sur les activités de l'entreprise qui les concernent, et ainsi d'agir sur le contexte.

- **Visualisation de données** : des visualisations intuitives aux analyses prédictives et au big data, TIBCO Spotfire® fournit instantanément aux utilisateurs les réponses qu'ils attendent et la possibilité de les partager avec leurs pairs, afin d'agir sur leurs découvertes.
- **Reporting et tableaux de bord** : les utilisateurs ont besoin d'accéder à l'analyse décisionnelle, au bon moment, au bon endroit, dans le bon contexte et sur le terminal ad-hoc. TIBCO JasperSoft® permet aux développeurs d'insuffler à toutes les applications la puissance de l'analyse décisionnelle.
- **L'analyse décisionnelle en streaming**. TIBCO Streambase® associe la connexion temps réel de vos activités aux informations découvertes par l'analyse décisionnelle afin d'identifier et d'agir sur les opportunités.



## AUGMENT INTELLIGENCE

### LA PLATEFORME DIGITALE DE TIBCO, TREMPLIN DES INITIATIVES NUMÉRIQUES IOT, MOBILE, CLOUD, API, BIG DATA

Avec la plateforme TIBCO, les données issues des objets, périphériques mobiles et autres systèmes sont combinées pour fournir une compréhension approfondie de chaque client en temps réel, ainsi qu'une expérience véritablement unique.

- **Initiatives internet des objets** : que ce soit pour des projets de traçabilité dans l'industrie ou les transports, pour offrir une expérience client plus riche et immersive avec de nouveaux services, la plateforme TIBCO ne se contente pas de vous connecter, elle donne à vos équipes la visibilité et la capacité requises pour agir et exploiter le plein potentiel de l'Internet des Objets.
- **Initiatives mobiles** : si la plateforme TIBCO assure des expériences incomparables via vos API, elle offre également l'analyse décisionnelle requise pour tirer parti de votre connaissance client et ainsi améliorer leur expérience.
- **Initiatives cloud** : la plateforme TIBCO prend en charge la modernisation des applications basées sur le cloud privé et les microservices. Elle permet l'adoption massive des solutions SaaS par les métiers. Elle nourrit la création d'API et garantit que l'entreprise dans son intégralité se présente comme une API évolutive afin de proposer ses services sur des plateformes externes, dont les réseaux sociaux.
- **Initiatives Big Data** : la plateforme TIBCO constitue la fondation idéale pour les initiatives Big Data, elle permet d'ingérer toutes les données, quels qu'en soient le volume, le débit ou le type. Elle se connecte de manière transparente aux solutions comme Hadoop et Spark. Facteur clé de différenciation de la plateforme TIBCO, l'analyse décisionnelle en streaming permet de mieux exploiter le Big Data et d'identifier instantanément les opportunités lorsqu'elles se présentent. La plateforme notifie alors automatiquement les ressources concernées ou génère la réponse appropriée, révélant ainsi toute la valeur des big data.



# TIBCO®

[www.tibco.com](http://www.tibco.com)

**INTERCONNECT  
EVERYTHING**



**AUGMENT  
INTELLIGENCE**

**Integration**

**API Management**

**Business Process  
Management**

**Data Visualization**

**Dashboard and Reporting**

**Streaming Analytics**

Pour toute information, veuillez contacter Jérôme Dardant au 01 44 51 45 88  
ou [jdardant@tibco.com](mailto:jdardant@tibco.com) - TIBCO Software - 25 rue Balzac 75008 Paris

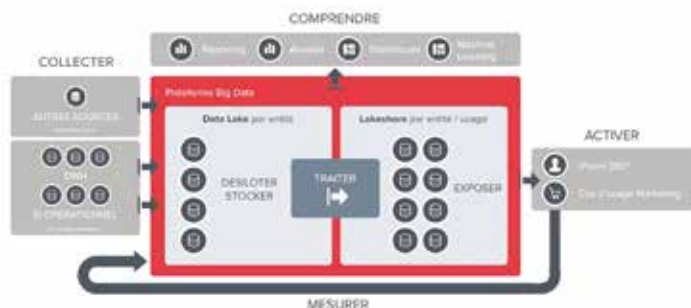


YSANCE / HARMONIE MUTUELLE : PROJET VISION 720

PORTEURS DU PROJET ET SES PARTENAIRES

Dans le cadre d'une série d'expérimentations de valorisation de ses données, utilisant les technologies de **Big Data**, **Harmonie Mutuelle** a choisi Ysance pour la réalisation d'un outil de **Vision 360°** à usage de la relation clients, avec un prolongement vers d'autres usages comme la segmentation client.

Le projet a été porté conjointement par les directions techniques et métiers : en particulier, la Direction Innovation Numérique, la Direction Architecture, Méthodes du GIE SIHM (qui gère les fonctions IT du groupe Harmonie Mutuelle) et la Direction Marketing d'Harmonie Mutuelle. Ysance, pour sa part, est intervenue avec une équipe polyvalente intégrant des expertises d'architecture et développement Big Data, data science et expérience utilisateur.



LE PROJET

Harmonie Mutuelle possède un **patrimoine de données** de grande valeur qu'elle gère actuellement dans une architecture organisée en **silos applicatifs**. L'accès et la modification des données se fait par le biais de logiciels différents selon les secteurs fonctionnels. Aussi, dans l'architecture de données actuelle, le rapprochement des données issues de secteurs fonctionnels différents et, par conséquence leur exploitation (y compris la Vision 360°), sont difficiles, voire impossibles à achever.

Démontrer la valeur des nouvelles technologies Big Data pour une entreprise comme Harmonie Mutuelle était le principal objectif du **Projet Vision 720**. Pour y arriver, nous avons choisi de bâtir un Data Lake à partir des données fournies par Harmonie Mutuelle et d'instancier plusieurs puits de données - ou « **LakeShores** » - en fonction des différents **cas d'usages**.

Les données brutes proviennent de quatre **silos de données** et de secteurs fonctionnels différents : référentiel personne/signalétique, activité commerciale/SAV/équipement, prestations et courriers de gestion sortants. Elles ont été chargées en l'état dans le système de fichiers distribués HDFS (Hadoop Files System) et comme tables externes dans Hive. Par la suite, une opération de **data crunching et de data cleaning** nous a permis de générer les tables Hive constituant le Data Lake.

Une série de cas d'usage pour prouver le potentiel d'une telle plateforme a été mise en place, pour lesquels plusieurs **LakeShores** ont été générés.

Tout d'abord, un moteur de recherche **Elasticsearch**, permettant d'effectuer des requêtes simples et complexes, a été construit pour permettre des **recherches rapides** dans l'**ensemble des données** à disposition.

Par la suite plusieurs cas d'usages relatifs à l'outil de visualisation ont été adressés, facilitant le rapprochement des données des différents silos et l'affichage sur une seule page des informations du client : informations personnelles, équipement, consommation et interactions avec la mutuelle. Des sources externes issus de l'Open Data (INSEE, IGN,...) ont permis d'enrichir à la fois les données et leur visualisation, d'où le nom de Vision 720 donné au projet.

Finalement, en capitalisant sur la plateforme Big Data et sa capacité à traiter de grands volumes, deux exemples d'études orientées marketing et employant des méthodes de **machine learning** ont été proposés :

- la réalisation d'un **clustering/segmentation** de la base des clients sans a priori, sur l'ensemble des données (et non sur des échantillons)
  - une étude se penchant sur les **facteurs déclenchant du SAV**.
- Ces études ont été menées sur des données entièrement anonymisées.

TECHNOLOGIES UTILISÉES

Les caractéristiques architecturales du projet Vision 720 ont été :

- La mise en place d'un **Data Lake** sur **Hadoop** (distribution Hortonworks), qui a permis de réconcilier facilement les données issues des différents silos et de mettre en place l'instrument de **visualisation clients**
- La construction de LakeShores avec Hive on Tez, permettant d'utiliser une simple syntaxe SQL-like (HiveQL) pour écrire les scripts d'alimentation qui néanmoins exploitent toute la puissance du **paradigme map-reduce** (et de son « évolution » Tez)
- La mise en place d'un moteur de recherche sur un serveur Elasticsearch permettant à l'utilisateur d'effectuer facilement des requêtes complexes sur plusieurs champs
- L'application de méthodes d'analyse classiques et modernes (**machine learning**) à des cas d'usages de « knowledge discovery » sur l'ensemble des données

VALORISATION OBSERVÉE DE LA DONNÉE

- Le data crunching met en évidence l'importance de la qualité des données.
- Un travail essentiel sur la réconciliation des données assure la pertinence du Data Lake.
- L'outil de visualisation permet aux utilisateurs un accès synthétique, rapide et transverse à l'ensemble des informations clients.
- La solution donne la possibilité d'effectuer des analyses de marché utilisant des approches modernes sur la totalité des clients.

ÉTAPES À VENIR ET DÉMARCHÉ BIG DATA CHEZ HARMONIE MUTUELLE

- L'expérimentation ayant montré une valeur métier, elle sera transformée en une solution pérenne
- Pour cela, la montée en compétence des équipes IT sera organisée au travers de l'internalisation de cette expérimentation
- D'autres expérimentations sont prévues afin d'aller plus loin dans les fonctions de « data visualisation »

# Ysance

YOUR DATA IN ACTION

## Nos offres



Data  
Integration



Data  
Architecture



Big Data



Data  
Science



Data  
Analytics

## Chaîne de valeur data



Compréhension et  
appropriation du  
métier



Analyse et  
définition de  
use cases



Traitement et  
enrichissement



Activation



Laissez vos données  
conter une autre histoire

**cloudera**