

NOUVELLES FRONTIÈRES ET NOUVEAUX ENJEUX  
DE LA GOUVERNANCE DE DONNÉES

## AVIS D'EXPERT

Par **Cyril Cohen-Solal** | Directeur de Keyrus Innovation Factory, **Keyrus**et **Xavier Dehan** | Directeur du Business Development Big Data & Analytics, **Keyrus**

Adoption toujours croissante du Cloud et des applications SaaS, généralisation de l'analytique embarqué, démocratisation des approches Big Data... Dans un contexte où les flux de données s'intensifient et où les activités métiers en sont de plus en plus dépendantes, la qualité et la cohérence des données diffusées deviennent primordiales. Aussi, à qui revient-il vraiment de s'en assurer ? En réarticulant les rôles des métiers et de l'IT, de nouveaux outils partagés permettent une gouvernance de données globale et durable<sup>1</sup>.

Selon des études récentes<sup>2</sup>, les entreprises estiment qu'en moyenne un tiers de leurs données sont inexactes. Elles sont 91% à penser que ces données erronées conduisent à des décisions inappropriées, affectant négativement leur performance économique. Enfin, 79% des organisations à travers le monde pensent que la majorité des décisions commerciales seront prises en fonction des données clients d'ici 2020. Si le constat n'est pas nouveau, la question de la responsabilité de la qualité des données reste entière et devient d'autant plus complexe que les données qu'il faudrait maîtriser se trouvent de plus en plus fréquemment à l'extérieur de l'entreprise, hors du contrôle direct des équipes IT. L'enjeu est critique, il s'agit en définitive, de constituer, d'enrichir, de préserver, et de partager, un capital de données de qualité : un actif essentiel pour valoriser l'entreprise, assurer son développement et garantir sa pérennité. Dès les années 1990, de nombreuses organisations ont engagé des stratégies de vérification et de validation afin de s'assurer que leurs applications sont « bien développées », qu'elles fonctionnent, et que l'on a réalisé « le bon produit » correspondant aux exigences utilisateurs. Les métiers et l'IT ont travaillé ensemble pour organiser ces activités, notamment en déployant des méthodes, des organisations et des outils de pilotage. A l'instar de la formule - en forme de constat, relative à la qualification logicielle - « les tests trop peu et trop tard ! », on trouve ici le principe « Garbage In – Garbage Out »<sup>3</sup>.

Il convient d'engager aujourd'hui la même démarche stratégique non plus pour s'assurer de la qualité des applications développées (du code et des spécifications) mais afin de maîtriser la qualité des données. Ces deux facteurs de risque (introduction d'anomalies logicielles et/ou de données erronées) contribuent à la non-qualité globale du SI de l'entreprise et génèrent des surcoûts (cycle de correction des défauts) et des baisses de performance (cycle analyses-décisions inappropriées).

## TROIS ÉVOLUTIONS QUI CHANGENT LE PÉRIMÈTRE DE LA GOUVERNANCE DE DONNÉES

• **Le succès des applications métiers en mode SaaS.**

L'offre croissante des éditeurs et les réductions de coûts associées aux solutions SaaS ont eu raison des réticences premières des entreprises à l'égard du Cloud. Par exemple, dans un domaine comme le CRM, sensible parce qu'il touche aux données clients, depuis 2014, les nouveaux déploiements se font majoritairement en mode SaaS et relèvent généralement de l'initiative des directions commerciales et/ou marketing. Le fait que les données des applications SaaS soient gérées selon des règles sur lesquelles l'entreprise n'a pas la main pose inévitablement aux départements IT des problèmes d'intégration et de cohérence avec les données des systèmes internes.

<sup>1</sup> Définitions de la Gouvernance des données :

• par le Data Governance Institute (DGI) : "Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods." [http://www.datagovernance.com/adg\\_data\\_governance\\_definition/](http://www.datagovernance.com/adg_data_governance_definition/)

• par le Cigref : « La gestion des données (ou Data Management) consiste à mettre en œuvre l'ensemble des dispositifs relatifs aux informations utilisées au sein de nos organisations afin d'en optimiser l'usage... La gouvernance des données en est le composant qui décrit les responsabilités, fixe les règles et contrôle leur application. Cette démarche de gouvernance est portée par une instance dédiée qui la pilote, avec des règles, des guides, des référentiels, des indicateurs, des chartes, une politique (de gestion des données personnelles, de classification, de stockage, de conservation, de brevets, de propriété intellectuelle, ...) »

« Comment gérer les données de l'entreprise pour créer de la valeur ? » Rapport Cigref – Enjeux Business des données <http://www.cigref.fr/wp/wp-content/uploads/2014/10/CIGREF-Enjeux-business-donnees-2014.pdf>

<sup>2</sup> «New Experian Data Quality research...» <http://www.experian.com/blogs/news/2015/01/29/data-quality-research-study/>

<sup>3</sup> « Data Quality Benchmark report 2015 » « Qualité et gestion des données – Tendances du marché en 2016 »

<sup>3</sup> GIGO : « données erronées en entrée, résultats erronés en sortie ».

• **Le développement de l'analytique embarqué et de l'in-Memory.** La plupart des applications métiers incorporent désormais des outils évolués d'analyse, de visualisation et de reporting qui prennent le pas sur les anciens systèmes de Business Intelligence centralisés. Ces derniers sont également concurrencés par les outils d'analyse in-Memory, moins complexes à mettre en œuvre et à utiliser. Si l'adoption de ces systèmes a l'avantage de rendre les équipes métiers beaucoup plus autonomes, elle se traduit également par une prolifération de référentiels dont personne n'est en mesure d'assurer la cohérence.

• **L'essor du Big Data.** Après des années de discours sur le Big Data, les entreprises ont enfin commencé à exploiter, en plus de leurs volumes croissants de données opérationnelles, les gigantesques sources de données multi-structurées que sont les réseaux sociaux et, plus largement, le social web. La maîtrise des règles de collecte et de transformation de ces données se révèle cruciale pour la pertinence des traitements algorithmiques (visions 360°, analyses prédictives, exigence temps réel,...) qui leur sont ensuite appliqués et conditionne la valeur des résultats obtenus. A ce jour, rares sont les entreprises à avoir étendu leur politique de qualité de données à ces nouveaux flux et processus.

La gouvernance unifiée des données constitue pour le Chief Data Officer, ce que représente la Théorie du Tout pour le physicien, tant ces 2 mondes de la Data semblent parfois dissociés !

Cette dissociation trouve ses causes principales dans l'essence même de la révolution Big Data avec sa dimension fortement expérimentale qui affecte de façon spécifique la gouvernance des données.

En effet, le Data Scientist travaille dans un *espace de liberté* afin de pouvoir tester en permanence son algorithme (modèle analytique et règles de gestion, architecture, écosystème de données) en intégrant à la volée des nouvelles sources de données (internes, externes, semi-structurées). Naturellement, cette activité est réalisée avec une traçabilité limitée, et le Data Scientist est plus focalisé sur les points de performances supplémentaires gagnés sur son cas d'usage métier ou technologique (réduction du Churn, amélioration du taux de prédiction, performance temps réel...) que sur l'alignement avec les canons de la qualité de données (exactitude, intégrité, unicité, conformité et complétude), les bonnes pratiques de la gestion des données, et l'impérieuse exigence d'industrialisation. Les entreprises répondent partiellement à cette problématique centrale avec des stratégies de gouvernance de données à minima, pour « border » les risques sur la qualité et la gestion des données, et en organisant des infrastructures comportant des espaces cloisonnés dédiés pour l'innovation, l'expérimentation, le durcissement de PoC et la pré-industrialisation (Lab Sandbox ...).

Aussi, les entreprises devront relever le défi majeur pour ces prochaines années que constitue l'organisation d'une gouvernance unifiée de leurs données : données de références, uniques, fiables, valides et intégrées (nom du client, adresse, produits, assets ...), données décisionnelles produites à partir des données transactionnelles (CA et marges par BU...) avec l'ensemble des données massives issues des infrastructures Big Data, Cloud et Digital qui comportent intrinsèquement un degré d'incertitude. C'est à cette condition, qu'elles pourront faire fructifier le capital des données qui constitue aujourd'hui un actif stratégique de l'entreprise pour créer de l'avantage concurrentiel, garantir le respect des exigences réglementaires et réduire les risques opérationnels sur l'ensemble des métiers de l'entreprises. Avec en corollaire à cette démarche, la convergence à termes des solutions Master Data Management et du Big Data.

Dans sa conclusion, l'étude menée par PAC pour le Syntec Numérique auprès de 100 organisations<sup>4</sup>, identifie quatre chantiers principaux pour accompagner la révolution Big Data dont la nécessité d'élaborer une stratégie de gouvernance de la donnée impliquant le DSI, les métiers, le Chief Data Officer... et définissant des règles d'accès et de sécurité de la donnée.

## UN BESOIN SOUS-ESTIMÉ : LA GESTION DES RÉFÉRENTIELS COMPLÉMENTAIRES

A ces évolutions, qui brouillent définitivement les frontières entre données internes et données externes, s'ajoute une problématique de plus en plus aiguë mais souvent sous-estimée : celle de l'absence de cohérence, de complétude des données métiers et des données référentielles qui les accompagnent (gestion de la qualité des données et des référentiels complémentaires). Avec l'automatisation d'une grande partie des processus de production de données, on peut considérer que l'immense majorité des données utilisées au sein d'une entreprise sont régies par des normes de qualité et font l'objet de contrôles en la matière.

Pourtant, dans de nombreuses entreprises, on observe qu'entre 2% et 5% des données sont gérées et collectées manuellement<sup>5</sup>, en dehors de toute application informatique pour répondre à des besoins ponctuels, transitoires, temporaires ou malheureusement durables de l'activité. Les fichiers Excel sont alors souvent les porteurs de ces données. Pour mesurer l'ampleur de cette problématique, il faut rapprocher cette réalité avec le constat que 56% des entreprises dans le monde déclarent que les erreurs de données proviennent majoritairement d'erreurs humaines<sup>2</sup>.

Trois exemples :

- Une société vient d'en racheter une autre. Chacune possède ses propres référentiels et, en attendant que les systèmes convergent, la Direction Financière crée un

<sup>4</sup> « De la data au Big Data, les attentes des utilisateurs métiers en France »

PAC pour le Syntec Numérique <http://www.syntec-numerique.fr/publication/data-au-big-data-attentes-utilisateurs-metiers-france>

<sup>5</sup> Etude interne Keyrus et analyse des projets clients

fichier Excel qui définit le mapping entre les référentiels des deux entités. Ce fichier géré manuellement est un référentiel complémentaire.

- Le référentiel « produits » n'étant pas encore intégré dans le système opérationnel du département marketing, ce dernier ne peut faire le lien entre une campagne et les produits qu'elle concerne. Pour chaque campagne, le département liste les produits concernés dans un fichier Excel et intègre ce fichier complémentaire dans son système.
- En l'absence de processus automatisé pour collecter les données de cashflow ou autres de ses filiales, le siège envoie chaque semaine un template Excel où chaque entité saisit manuellement ses données. Cela se pratique dans les plus grandes entreprises !

Ces données et référentiels complémentaires gérés en dehors des applications informatiques échappent à tout processus de contrôle rigoureux au niveau :

- 1/ **de la saisie des données**, d'où des erreurs qui, en se propageant, conduisent à des incohérences dans les applications aval ;
- 2/ **de la traçabilité de la collecte elle-même**, Excel ne permettant pas de suivre avec certitude des critères tels que les dates effectives de mise à jour des données, ni de savoir si toutes les personnes concernées ont bien reçu le formulaire de collecte ;
- 3/ **de la qualité de l'intégration des fichiers avec les systèmes opérationnels**, les correspondances pouvant être brisées par l'ajout/la suppression d'un élément ou simplement parce qu'un nom n'est pas orthographié de la même façon.

## PERMETTRE AUX MÉTIERS ET À L'IT D'ÊTRE ACTEURS DE LA GOUVERNANCE DE DONNÉES

Les évolutions et les pratiques qui viennent d'être décrites obligent les entreprises à abandonner les politiques de qualité de données en silo ou limitées au seul système d'information décisionnel. Elles plaident en faveur d'une gouvernance de données globale permettant aux métiers comme aux départements IT de contribuer à la qualité de données globale de l'entreprise. Sachant que l'absence d'outil simple est un obstacle majeur à l'inscription d'une telle politique dans la durée, Keyrus propose une approche adossée à une plateforme de Data Governance partagée par les équipes IT et les différentes équipes métiers. Cette approche réaffirme et réarticule les rôles respectifs de ces équipes :

- L'IT, garant de la cohérence globale du système d'information étendu de l'entreprise, doit avoir la main sur les problématiques de qualité technique des données. Typiquement, c'est à lui de mettre en place les tests appropriés pour s'assurer de l'exacte correspondance

des données entre un système interne et une application Cloud, de veiller à l'absence de doublons, à l'intégrité des sources, à la continuité des flux de données, etc.

- Les équipes Métiers, en tant qu'utilisatrices des données, doivent avoir la main sur la qualité fonctionnelle des données. Elles sont en effet mieux placées que les équipes IT pour détecter des incohérences au niveau des indicateurs qu'elles utilisent au quotidien ou des incohérences par rapport à un historique, ainsi que pour définir des contrôles tels que des seuils de pertinence et des règles de vérification.

Une plateforme de Data Governance telle que **quillup**<sup>6</sup> permet à ces deux catégories d'acteurs de contrôler et surveiller la qualité ainsi que la cohérence des données, quelle que soit la nature des sources (base de données, application, fichier, cubes, etc.) dès lors que ces dernières ont été déclarées sur la plateforme et que les autorisations d'accès ont été attribuées.

## ÉLIMINER LES POINTS DE FAIBLESSE DE LA GOUVERNANCE

La qualité de données n'est jamais acquise. La collecte de données complémentaires et la création de nouveaux référentiels sont inévitables dans la vie d'une organisation. Ce sont aussi, du fait de l'absence de contrôles, autant de points faibles dans un dispositif de gouvernance de données à l'échelle de l'entreprise. Inclure dans la plateforme de Data Governance partagée par l'IT et les métiers un outil spécifiquement dédié à la gestion des données et référentiels complémentaires permet d'éliminer le problème à la source. Au lieu d'utiliser Excel pour collecter les données devant alimenter un système aval, les utilisateurs métiers peuvent le faire de manière rigoureuse et structurée via des formulaires bénéficiant automatiquement de contrôles de saisie et de cohérence. Cela revient à éliminer le recours aux fichiers Excel et les erreurs qu'ils sont susceptibles de propager dans les chaînes opérationnelles et décisionnelles. Donner cette possibilité aux équipes métiers, c'est résorber les 2 à 5% de données échappant à tout contrôle de qualité et de cohérence sans pour autant faire intervenir les équipes IT. C'est aussi le moyen de limiter les risques d'incohérence et les pertes de traçabilité lors des périodes de transition où il faut faire le pont entre deux systèmes (fusion, migration, etc.) et maintenir des tables de correspondance et des référentiels temporaires.

## A LA CLÉ : DES GAINS D'EFFICACITÉ ET UNE CONFIANCE RETROUVÉE DANS LES DONNÉES

Les entreprises qui ont opté pour cette approche et qui s'appuient sur la plateforme **quillup** pour assurer la gouvernance de leurs données estiment à 30% le gain sur les coûts de maintenance pour le service IT, et à 80% le gain de temps pour les utilisateurs métiers, qui n'ont plus à effectuer de contrôles manuels, ni à solliciter l'IT. On peut

<sup>6</sup> quillup : plate-forme haute performance de gouvernance des données conçue pour aider les entreprises à améliorer la qualité de leurs données et à fiabiliser leurs prises de décisions, développée par Keyrus au travers de son Centre de Recherche & Développement basé en Israël. Pour en savoir plus : <http://quillup.com/>  
Livres de référence : « La qualité et la gouvernance des données au service de la performance des entreprises », de Laure Berti-Equille - éditions Hermès - Lavoisier  
"Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program", de John Ladley - éditions Morgan Kaufmann



souligner que, si l'objectif est bien de couvrir l'ensemble des données de l'entreprise, la mise en place d'une telle gouvernance doit se faire de manière progressive, typiquement en commençant par un département métier pilote.

La condition sine qua non pour réussir est évidemment d'impliquer dans le projet les deux catégories d'acteurs. Du côté des départements IT, les éventuelles objections à une gouvernance partagée disparaissent dès lors qu'ils sont assurés de conserver la maîtrise de la qualité technique des données et des accès aux différents systèmes. Les utilisateurs métiers, contrairement à ce qu'on imagine, sont loin d'être réticents à l'idée de s'impliquer dans la gestion de la qualité de leurs données. Ils perçoivent

rapidement ce qu'ils ont à y gagner en termes d'autonomie, d'efficacité et, surtout, de confiance dans les données qu'ils utilisent au quotidien à des fins opérationnelles ou de prise de décisions.

Dans un environnement où les données seront de plus en plus issues de sources tierces ne dépendant pas de l'entreprise (web, réseaux sociaux, objets connectés, etc.), ce dernier point est essentiel : la confiance des opérationnels dans la qualité des données conditionne directement la valeur qu'ils seront en mesure de créer à partir de ces données. Au-delà des bénéfices immédiats que sont les gains d'efficacité et les réductions de coûts, là est sans doute le véritable enjeu d'une gouvernance globale des données.

C.C.S.

X.D.

#### A PROPOS DES AUTEURS

Diplômé d'Informatique de Gestion (Paris Dauphine), **Cyril Cohen-Solal** possède une expertise de 15 années dans le domaine de la Data Intelligence qui lui a permis d'accompagner les plus grandes entreprises en France et à l'International dans la conception et la mise en œuvre de leur stratégie décisionnelle. **Cyril** est également en charge, au sein du **Groupe Keyrus**, de notre accélérateur « Keyrus Innovation Factory », passerelle entre les start-ups innovantes et les grandes entreprises européennes. En parallèle à ces activités, il enseigne depuis 10 ans à Paris Dauphine dans le cadre du Master Informatique Décisionnel.

**Xavier Dehan** est Directeur du Business Development Big Data & Analytics au sein du **Groupe Keyrus**. Il possède une large expérience dans le domaine de l'intégration de projet IT et la Qualité Logicielle, acquise auprès de grandes Sociétés de Services et dans le cadre de la création de sociétés Pure Players Quality Insurance & Testing. Depuis plus de 10 ans, il conseille les grandes organisations sur leur stratégie [Big] Data pour la réalisation de projets expérimentaux et industriels sur toute la chaîne de valorisation des données.

#### A PROPOS DE KEYRUS

##### 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, **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 :

- **Data Intelligence**

Data Science - Big Data Analytics – Business Intelligence – EIM – CPM/EPM

- **Digital Experience**

Innovation & Stratégie Digitale – Marketing Digital & CRM – Commerce Digital – Performance Digitale – User Experience

- **Conseil en Management & Transformation**

Stratégie & Innovation – Transformation Digitale – Pilotage de la Performance – Accompagnement des Projets

Présent dans 15 pays et sur 4 continents, le Groupe **Keyrus** emploie 2500 collaborateurs.

**Keyrus** est coté sur le compartiment C de l'Eurolist d'Euronext Paris (Compartiment C/Small caps - Code ISIN : FR0004029411 – Reuters : KEYR.PA – Bloomberg : KEY:FP)

Plus d'informations sur : [www.keyrus.fr](http://www.keyrus.fr)