MYRIAD
ACTIONABLE DATA

La gouvernance de la donnée : condition de la réussite de vos projets IA

Proposé par
Charles-Eric de La Chapelle

Sommaire

Introduction

L'intelligence Artificielle, en tant que partie prenante du processus de digitalisation de l'entreprise, connaît un essor impressionnant sur un grand nombre de secteurs d'activité et ses cas d'applications se multiplient. Les entreprises s'appuient de plus en plus sur l'analyse des données et leur exploitation, pour adapter leur stratégie, prédire les différents scénarios possibles, automatiser les processus et réduire leurs coûts.

La donnée est ainsi devenue le carburant du système d'information. Elle est aujourd'hui interne comme externe, structurée comme non structurée, produite par les humains ou des objets connectés. Pour améliorer l'information, ouvrir de nouvelles opportunités et réduire les coûts les DSI doivent garantir à l'entreprise que les données traitées répondent aux exigences définies par le CobiT* : efficacité, efficience, confidentialité, intégrité, disponibilité, conformité et fiabilité. Ces données sont un capital dont la volumétrie ne va pas cesser de croître (Cf. tableau ci-dessous) notamment avec le développement continu de l'IoT (Internet of Things). Ceci implique l'acquisition de compétences nouvelles, notamment en matière de réseaux, d'outils de traitement et de connaissance des méthodes de mise en place de modèles de calculs (data sciences & analytics).

Dans ce contexte il est donc nécessaire d'établir un cadre d'organisation permettant d'établir la stratégie, les objectifs et les politiques permettant de gérer efficacement les données de l'entreprise. Faute d'établir ce cadre, de nombreuses entreprises ont vu leur projet de développement ou d'amélioration de leur compétitivité remis en cause que ce soit par leurs salariés, désorientés par l'inefficacité des mesures mises en place et ne contribuant pas au système, et/ou sanctionnées par leurs clients mécontents de la qualité ou de la performance du service apporté.

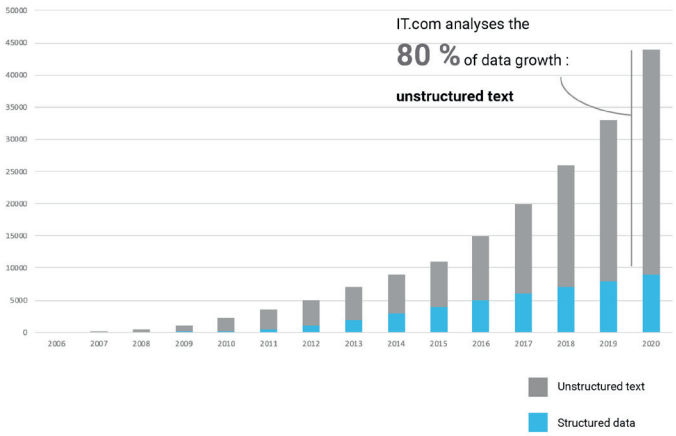
C'est ce cadre, appelé gouvernance, dont nous allons tenter dans ce livre blanc de définir la nature et les contours, qui en sont les acteurs, ce que sa mise en place implique et présenter plusieurs cas d'applications :

- Cas 1 : Mettre en place une gouvernance de la donnée en partant des exigences réglementaires autour de la qualité des données
- Cas 2 : Mettre en place une gouvernance de la donnée afin de permettre à 5 marques de voiture d'utiliser une plateforme digitale commune pour leurs 270 sites internet
- Cas 3 : Normalisation de la donnée digitale/comportementale pour la mise en place du projet Client 360°

Ces cas d'applications montrent à quel point une bonne gouvernance de la donnée est nécessaire pour valoriser des données souvent peu utilisées et réussir les programmes impliquant notamment de l'intelligence artificielle.

*le COBIT est un cadre de référence pour maitriser la gouvernance des SI dans le temps. Il est fondé sur un ensemble de bonnes pratiques collectées auprès d'experts du SI.

Pour prendre un exemple et donner un ordre de grandeur de l'importance de la croissance de la donnée, et en ne reprenant uniquement que les données textuelles, nous pouvons voir la projection ci-dessous de la croissance mondiale de la donnée de l'entreprise (Source : IDC, The Digital Universe 2010)



La donnée, qu'est-ce que c'est ?

Par définition, une donnée est **un élément brut**, qui n'a pas encore été interprété, ni mis en contexte.

Et c'est là toute la différence entre une information et une donnée. En effet, **une information est une donnée interprétée**. En d'autres termes, la mise en contexte d'une donnée crée de la valeur ajoutée pour constituer une information. Une donnée est un élément défini et isolable qui va pouvoir être manipulé, traité et analysé en fonction d'un objectif ou d'un cadre d'analyse.

L'information naît donc de la relation entre une donnée et une personne ou un modèle algorithmique, qui lui confère du sens.

En d'autres termes, les données sont des éléments de base qui en fonction du contexte seront comprises et traduites par des hommes ou des machines. Comprendre les données et les constituer en informations **constitue la base** de la culture de la donnée.

Prenons l'exemple de ces balles de golf écrit par Kaptain Kobold sur Flickr. Que pouvons-nous en dire ? Ce sont des balles de golf. L'une des premières choses que l'on sait est donc qu'elles sont utilisées pour jouer... au golf. Par ailleurs, le golf est un sport, ce qui nous permet de placer la balle de golf dans une taxonomie. Même les objets d'apparence banale recèlent en réalité une quantité de données importantes qui leurs sont attachées. Vous aussi, vous avez un nom de famille, une date de naissance, un poids, une taille, une nationalité, etc. Toutes ces choses sont des données.

Dans l'exemple ci-dessus, nous pouvons déjà constater qu'il y a différents types de données. Il y a principalement des **données qualitatives** et des **données quantitatives**.

- **Les données qualitatives** se réfèrent à la qualité : La description d'une couleur, de textures et l'aspect d'un objet, la description d'une expérience sont toutes des données qualitatives.
- **Les données quantitatives** sont des données qui se réfèrent aux chiffres. Ex : Le nombre de balles de golf, la taille, le prix, le résultat d'un test, etc.

Cependant, vous allez rencontrer d'autres types de données :

- **Les données catégorielles** permettent de classer les objets que vous traitez par catégories. Dans notre exemple, l'aspect « usagé » serait une catégorie au sein de la typologie suivante : « nouveau », « usagé », « cassé », etc.
- **Les données discrètes** sont des données dénombrables. Ex : le nombre de balles de golf. Il ne peut y avoir qu'un nombre entier de balles de golf (il ne peut pas y avoir 0,3 balles de golf). Le résultat d'un test ou une pointure de chaussure constituent d'autres exemples.
- **Les données continues** sont des données numériques non entières. Ex: le diamètre des balles de golf (ex: 10,53mm, 10,56mm, 10.536mm), ou la taille précise de votre pied (en opposition à la pointure, qui elle est discrète). Toutes les valeurs sont admises.

De la donnée, à l'information, à la connaissance

Les données, quand elles sont collectées et structurées deviennent soudain très utiles. Structurons les dans le tableau ci-dessous :

Couleur	Blanche
Catégories	Sport, Golf
État	Usagé
Diamètre	43mm
Prix (par balle)	0,36 €

Ces données n'ont pas de sens exploitées individuellement. Pour faire émerger l'information, nous devons les interpréter.

Prenons la taille : Un diamètre de 43 mm ne signifie rien. Il devient intéressant quand il est comparé à une autre donnée, un autre diamètre. Dans certains sports, il y a une réglementation pour les équipements. La taille minimale d'une balle de golf en compétition est de 42,67 mm. Nous pouvons donc utiliser cette balle en compétition. C'est une information. En revanche, ce n'est toujours pas de la connaissance. La connaissance est créée lorsque l'information est apprise, appliquée et comprise.

Données non structurées contre données structurées

- **Les données interprétables**

« Il y a 5 balles de golf usagées avec un diamètre de 43 mm à 0,5 € chacune » est une phrase facilement compréhensible pour un humain, mais compliquée à comprendre par un ordinateur. La phrase ci-dessus est considérée comme de la donnée non structurée. Elle n'a pas de structure sous-jacente. La tournure de la phrase pourrait être changée et il n'est pas évident de déterminer quel mot correspond à quelle donnée. De la même manière, les PDFs et les images peuvent contenir des informations interprétables par l'œil humain, mais ne pas être toujours compréhensibles par un ordinateur.

- **Les données interprétables par l'ordinateur**

Si l'on veut que l'ordinateur analyse la donnée, il faut qu'il soit capable de la lire et de la traiter. Ce qui signifie qu'elle doit être structurée dans un format lisible par la machine.

Définition

La gouvernance des données est le cadre d'organisation permettant d'établir la stratégie, les objectifs et les politiques dont la résultante est une gestion efficace des données de l'entreprise. Il comprend les processus, les politiques, l'organisation, les compétences et les technologies nécessaires pour gérer et garantir la disponibilité, la facilité d'utilisation, l'intégrité, la constance, l'auditabilité et la sécurité de vos données.

Qu'est-ce que la gouvernance ?

La gouvernance des données était initialement définie comme une tâche informatique.

Chaque projet métier mettait en œuvre une gouvernance de la donnée dédiée. Un cadre était utilisé qui créait une structure de données figée, les données, quand elles étaient accessibles et utilisables, étant formatées en fonction de l'usage qui en était fait, tout ceci en assurant la sécurité et la confidentialité des données de manière efficace.

Cette façon de procéder créait des silos, elle ne permettait pas, dans une même organisation, de profiter des opportunités que pouvaient offrir le partage des données et encore moins d'utiliser des données externes à l'entreprise.

Les nouvelles opportunités qu'ont apporté Internet comme source d'information, le Big Data comme vecteur de rapidité d'utilisation de données massives et variées, ainsi que les exigences en constante évolution définies par la réglementation ont poussé les entreprises à aller au-delà. Désormais, les entreprises adoptent une approche plus unifiée de gestion pour extraire de la valeur de ces données et en assurer la gestion.

Ainsi est apparu une nouvelle vision que l'essor d'internet, des technologies-en particulier cloud et Big Data ont grandement favorisé et qui apporte beaucoup plus de valeur aux données que les entreprises collectent, maintiennent et protègent.

La gouvernance des données consiste à gérer les données pour les normaliser et à harmoniser la manière dont elles sont utilisées pour les initiatives des différents métiers de l'entreprise. La gouvernance des données permet ainsi de garantir que les données critiques sont disponibles au bon moment, pour la bonne personne, dans une forme standardisée et fiable. Un avantage qui favorise une meilleure organisation des opérations et se traduit par une productivité et une efficacité accrue.

Une bonne gouvernance des données induit des pratiques qui optimisent la valeur des données en permettant une propriété et une traçabilité complète de la vie de la donnée. Les entreprises peuvent prendre des décisions en s'appuyant sur une donnée fiable et vraie. Ainsi, La gouvernance de la donnée devient la clef de voute d'un système de gestion d'entreprise dans lequel tous les acteurs, à tous les niveaux, sont impliqués de manière formelle, responsable, proactive et efficace pour garantir la confiance et l'efficacité.

Mémo de Jeff Bezos
(révélé en 2011 par une fuite de Steve Yegge : https://frama.link/BezosLeak) Date : 2002

Collaborateurs concernés : TOUS

Contenu :

1. Toutes les équipes devront désormais exposer leurs données et leurs fonctionnalités par des interfaces de services à travers le réseau (API).
2. Les équipes doivent impérativement communiquer entre elles à travers des interfaces.
3. Il n'y aura aucune autre forme de communication permise : pas de liens directs, pas de lecture directe de l'entrepôt de données d'une autre équipe, pas de lecture en mémoire partagée, pas de portes dérobées ou autres. La seule communication permise est à travers des appels d'interface de service à travers le réseau.
4. La technologie n'a pas d'importance. HTTP, Corba, Pubsub, protocoles spécifiques...
5. Toutes les interfaces de service, sans exception, doivent être conçues pour être externalisables. L'équipe doit planifier et concevoir l'interface afin de pouvoir l'exposer aux développeurs du monde extérieur. Sans exception.
6. Quiconque ne respectera pas ces principes sera viré. (librement traduit par Charles Népote)

Conséquences sur la structure de l'entreprise :

- Multiplication des services proposés via des interfaces : temps machine (EC2), stockage classique (S3)...
- Création d'une culture de l'exposition et de circulation des données
- Constitution d'Amazon en une entreprise plateforme de données accessibles par des mécanismes simples et documentés

Source : infolabs

Faut-il une culture et une transversalité de la donnée ?

Il est aujourd'hui impossible pour les entreprises et les organisations d'échapper à la donnée. Elle s'invite partout dans les projets, dans les fiches de poste, dans les relations avec les fournisseurs qui ont construit leur « business model » autour du traitement de la donnée, et même la règlementation s'en mêle avec loi pour la république numérique en France ou le RGPD en Europe. Pourtant, en dépit de son volume sans cesse grandissant, la plupart du temps **peu de collaborateurs** ont les compétences pour les traiter et en tirer de la valeur pour l'entreprise, au-delà des usages les plus banalisés, à travers des outils CRM, de suivi de la production, ou de reporting financier.

Il convient donc **de développer une culture de la donnée** qui permette de communiquer de manière critique et d'appréhender comment tenir compte des résultats du traitement des données lors de discussions ou de prises de décisions par exemple.

Les GAFA en standardisant des modèles économiques et des offres reposant sur la donnée ont créé de nouvelles attentes des consommateurs et des clients. Cela a été rendu possible parce qu'elles ont la placé la culture de la donnée **à tous les niveaux de leurs entreprises.**

Dans ce contexte, il est impératif d'embarquer tous les salariés à tous les niveaux pour développer une culture « data centric ».

Nous allons essayer de définir quelques pistes permettant d'acquérir et de transversaliser la culture de la donnée.
Au départ il faut définir la donnée afin que tous les acteurs comprennent bien la séparation entre donnée et information Cf chapitre : « En fait QU'EST-CE QUE LA DONNEE ? ». Ensuite, il faut organiser la mise à disposition des données en interne à travers une plateforme partagée, comme un datalake et encourager par tous les moyens possible les collaborateurs à **s'emparer de ce patrimoine** de donnée et à comprendre la nécessité d'en devenir **responsables** pour créer les **bons réflexes**.

Dans cette phase, l'entreprise va encourager des

actions d'**expérimentation** sur les données en combinant ces actions avec **la mise en place de règles de gestion et d'utilisation des données**. Dans le même temps doit être entrepris une **sensibilisation des acteurs sur les causes de corruption, de perte et de vol des données.**

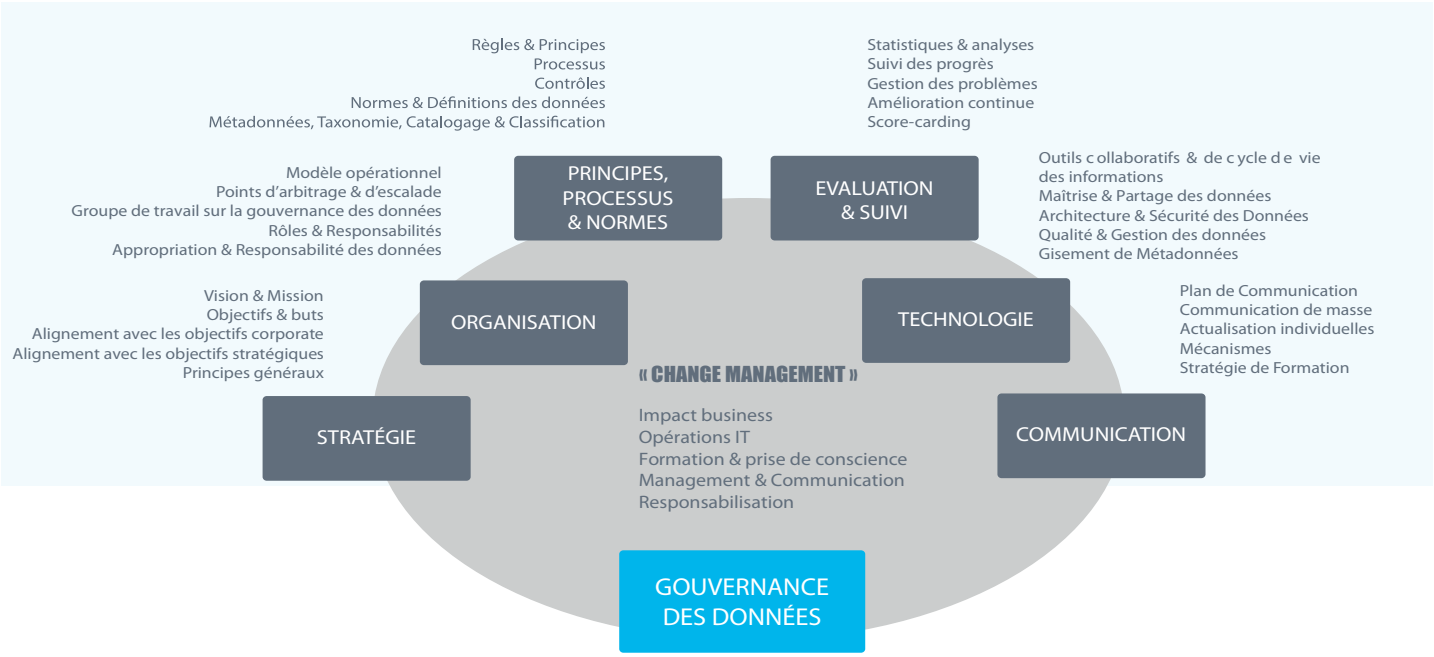
Parmi ces actions concrètes visant à faire prendre conscience et à utiliser le patrimoine de donnée on peut citer :

- La **cartographie des données**, qui consiste à identifier les principales sources de données pertinentes à valoriser, leur producteurs et les flux associés
- La mise en œuvre de **« Quick Wins »** Il s'agit de promouvoir des actions permettant la mise en œuvre par les métiers de cas d'usage sur les données ayant un retour sur investissement (ROI) rapide et visible par toute l'organisation
- La mise en place d'**une politique de Gouvernance, de Sécurité et de confidentialité des données** qui pourra se traduire par exemple dans une charte largement partagée
- Le choix et l'intégration de solutions (de traitement) et d'échanges sur les meilleures pratiques.

Développer et mettre en place une stratégie d'entreprise orientée sur les données suppose :

- Prise de conscience
- Engagement à tous les niveaux
- Mise à disposition des moyens adéquats

Cadre de programme de gouvernance des données



Une vision qui implique toute l'organisation ?

L'entreprise ou l'organisation ayant pris conscience de la valeur stratégique des données et des usages qui peuvent en découler, il s'agit maintenant **de former et de responsabiliser** les producteurs de donnée sur l'importance cruciale de leur apport dans le système d'information (Commerciaux qui saisissent dans le CRM, rapport de maintenance dans l'industrie, saisie comptable ou RH etc…) En parallèle à cette implication nécessaire des acteurs historiques de l'organisation dans le processus de création et de cycle de vie de la donnée, **de nouveaux métiers dédiés** sont devenus nécessaires pour garantir le traitement des données.

En haut de la pyramide de ces nouveaux métiers on trouve le **Chief Data Officer (CDO)**, souvent rattaché à la direction Générale, chargé de **piloter les actions métiers autour de la donnée** et de faire avancer l'entreprise dans sa transformation digitale par la donnée. Au-delà du CDO, bien d'autres métiers sont apparus comme le **Chef de projet Data** (ou Data Manager), le **Data Engineer** (qui développe et entretient les systèmes de collecte, stockage et mise à disposition des données), le **Data Scientist** (qui récupère, traite, gère et analyse des données, élabore des modèles prédictifs et développe des algorithmes), ou encore d'autres, plus fonctionnels, ou liés à l'évolution des usages et de la législation, comme les Data Privacy Officer (DPO).

Dans le tableau ci-après vous verrez un exemple de profils qui peuvent être amenés à travailler dans le cadre d'un programme de transformation numérique centré sur la donnée.

Expertise	Famille de métier	Profil	Définition profil
CONSEIL	Management de projet	Chef de projet data / Data Product Manager	Le Chef de projet data assure la gestion de projets (Big) data. Il étudie la faisabilité du projet, optimise la répartition des ressources humaines et techniques en conformité avec le cahier des charges et veille au respect des délais, des coûts, des performances et de la qualité du produit réalisé en appui du Product Owner.
DATA	Analyse et stratégie data	Product Owner	Un Product Owner identifie les besoins au sein d'une organisation dans le domaine data. Il est également chargé de déterminer un plan d'action approprié ou une solution en fonction des besoins des métiers et d'animer l'offre correspondante.
DATA	Analyse et stratégie data	Data Strategist	Le Data Strategist réalise pour l'entreprise des études stratégiques dans le domaine data (compréhension des enjeux, détermination des grandes orientations, monétisation des données). Expert en management et en organisation des données, il assure pour l'entreprise des missions d'audit et d'analyse de l'existant (algorithmes, plateforme, sécurité).
DATA	Analyse et valorisation des données	Data Scientist	Le Data Scientist est responsable de la gestion et de l'analyse des données (dans un contexte big data ou non). Il est en charge de la récupération, du traitement de la donnée, de l'élaboration de modèles prédictifs et de la construction des algorithmes, puis du test des résultats.

Expertise	Famille de métier	Profil	Définition profil
DATA	Analyse et valorisation des données	Dataminer	Le Data miner valorise l'ensemble des données client pour en faire un levier de création de valeur pour l'entreprise. Il analyse des masses de données hétérogènes, éventuellement non structurées, pour en extraire de la connaissance utile à l'optimisation des offres et services de l'entreprise.
DATA	Analyse et valorisation des données	Formateur Data Science	Le Formateur Data Science a un profil de Data Scientist Senior. Il intervient dans une équipe Data Science d'une entreprise pour assurer la formation théorique et pratique sur les sujets de Data Science intéressant l'équipe, ainsi que pour du tutoring et de la montée en compétences sur des projets concrets.
DATA	Architecture et modélisation des données	Ingénieur Big Data	L'Ingénieur Big Data est chargé de l'automatisation du fonctionnement de la plateforme Big Data, ainsi que de la partie mise en production / mise à l'échelle dans un contexte Big Data des traitements et modèles prédictifs.
DATA	Architecture et modélisation des données	Architecte urbaniste	L'Architecte urbaniste est chargé de l'audit et des préconisations fonctionnelles du Système d'Information du client. Il intervient notamment sur les aspects de modélisation, de la qualité et traçabilité des données, et des problématiques propres à l'interaction entre les environnements Big Data et le SI du client.
DATA	Architecture et modélisation des données / Analyse et valorisation des données	Data Analyst	Le Data Analyst crée les bases de données nécessaires à l'entreprise puis s'assure de leur bon fonctionnement. Ainsi il gère l'administration et l'architecture des bases mais est aussi en charge de la modélisation des données. Disposant aussi d'une connaissance métier du secteur de l'entreprise, il a pour mission d'élaborer des critères de segmentation afin d'étudier au mieux les chiffres et participe à l'analyse et à l'exploration des données.
DATA	Data Manager	Data manager	Le Data Manager ou gestionnaire de données, est en charge d'acquérir et d'organiser l'ensemble des données de l'entreprise dont il a la charge pour rendre leur exploitation optimale. Il doit être à l'aise pour comprendre la base de données qu'il structure, et faire en sorte qu'elle soit compréhensible.
GOUVERNANCE / QD	Gestion et administration	Architecte Big Data	L'Architecte Big Data est chargé des études et préconisations techniques de l'architecture d'une plateforme Big Data. Puis il gère la mise en place de la plateforme, ses interactions avec les sources de données, le SI du client et les partenaires externes.

Expertise	Famille de métier	Profil	Définition profil
CONSEIL	Gouvernance et qualité des données	Data Steward	Le Data Steward est un garant de la qualité de la donnée. Son rôle est de capturer l'information, de documenter la donnée avec un certain nombre d'éléments de définition et de cartographie, et de valider le niveau de qualité global des données. Sur le Datalake, le Data Steward contribue au nommage métier, au traitement des doublons, veille à la bonne gestion du cycle de vie et à la traçabilité des données.
CONSEIL	Management de projet	Directeur de projet data / Data Product Manager Senior	Le Data Product Manager Senior a pour mission de concevoir et mettre en place les projets data de l'entreprise : il coordonne les équipes fonctionnelles et techniques, suit les budgets et la réalisation des projets en appui du Product Owner. Il dispose d'une expérience importante sur les projets data.
CONSEIL /DATA	Visualisation des données	Consultant Dataviz	Le Consultant Dataviz a pour but de présenter les données sous forme visuelle (reporting, interfaces, graphiques) afin d'en faciliter la compréhension et / ou l'analyse.

Gouvernance et plateforme unifiée de données

Traditionnellement, dans le passé, on utilisait des méthodologies, des processus et des outils distincts pour gérer les données structurées et les données non structurées. Le plus souvent les processus et les outils étaient tous pris en charge par les services informatiques, qu'il s'agisse du stockage, de la sécurité, comme des outils métiers. Le plus souvent, la gouvernance pratiquée par les DSI servait à garantir que les données répondent aux exigences techniques et de compliance des projets et aidait les entreprises à adresser leurs besoins métiers en évitant les problèmes juridiques ou financiers liés à des violations de conformité.

Pour cette raison, les **procédures** (mais aussi, pour des aléas techniques, les structures) utilisées pour le traitement de la donnée **étaient rigides**. Cela permettait effectivement de répondre aux exigences de sécurité, de robustesse et de conformité, ainsi qu'à un besoin de prises en charges des outils des métiers. Dans ces conditions, un partage des données **n'était pas réellement possible** et les **capacités d'innovation des entreprises s'en trouvaient freinées**. Également, l'analyse des données était fastidieuse et gourmande en temps comme en ressources.

Aujourd'hui, les nouvelles organisations mises en place reposent sur **une gestion unifiée des ressources et un accès facilité aux données** et permettent aux entreprises non seulement de mieux s'adapter à leur environnement mais aussi d'innover.

Il s'agit bien, pour l'entreprise qui parle de gouvernance, d'avoir une vision très claire que ses données constituent un **actif stratégique**. C'est cette vision qui **motive** et qui est le **moteur du changement**. En effet, au-delà de la gouvernance, c'est **l'approche unifiée de la gestion des données structurées et non structurées** qui devient la clé pour gérer la croissance exponentielle des données, générée par la numérisation des processus et des outils (l'IoT sera également à l'origine de volumes importants de données semi-structurées)

Il peut être compliqué d'organiser les données et de les regrouper tout en facilitant l'accès puis l'analyse au sein d'une plateforme unifiée de gestion des données. Les écueils rencontrés ne proviennent pas nécessairement de l'architecture physique ou d'un nouvel environnement (tel que Hadoop par exemple) mais concernent davantage des lacunes de gestion des données.

Les étapes suivantes, non exclusives, peuvent surmonter ces problèmes et assurer la mise en place de la gouvernance de la plateforme :

1. **Classer les données** : l'organisation des données dans la plateforme demande de les classer (type, contenu, origine, utilisation(s), utilisateur(s), caractère personnel etc) ;
2. **Identifier une architecture cible** : la classification préalablement établie permet d'organiser les données dans la plateforme et par conséquent d'y définir les schémas d'ingestion/migration de ces données.
3. **Définir une architecture appropriée** : la hiérarchie des fichiers pour le stockage des données et la mise au point de conventions de nommage des fichiers/répertoires sont des points clés pour la gestion des données dans la plateforme.
4. **Assurer une architecture fluide et sécurisée** : les mécanismes de contrôle, d'accès et de distribution des données en fonction des utilisations/utilisateurs (création d'un catalogue de données) sont également à mettre au point en vue d'assurer la bonne circulation des données. En outre, la maîtrise de la qualité des données sera un facteur clé de succès de l'efficacité fonctionnelle de la plateforme.

Pour mettre en œuvre ce concept, tous les facteurs, **personnes, stratégies et processus doivent être combinés** dans le but d'assurer **l'efficacité organisationnelle** avec **des moyens identifiés** et **des indicateurs clairs** pour mesurer le succès. L'idée est donc de que toute l'entreprise soit alignée, **en incluant les services informatiques**, dans l'objectif de partager et d'étendre la valeur des données en les enrichissant, afin d'augmenter la connaissance pour stimuler la productivité et l'innovation.

Comment mettre en place les bons indicateurs ?

La mise en place d'indicateurs pertinents et d'indicateurs de performance (KPI) est essentielle à la réussite d'un programme de gouvernance afin de bien aligner la pertinence des actions mises en œuvre au regard de la valeur attendue en sortie.

Les indicateurs peuvent être de toute nature, par exemple :

- Nombre de demandes (d'accès à des données) des métiers enregistrés
- Nombre de propriétaires de données (Data Owner) identifiés
- % des demandes métiers adressées en regard des SLA mis en place…

Les KPI sont des indicateurs choisis par la direction générale pour mesurer la performance du programme de gouvernance. Les KPI sont utilisés pour le pilotage des projets, apporter des améliorations ou mener des actions correctives. Ils reflètent les facteurs clés de succès du programme.

Nous vous proposons de diviser en 4 catégories les indicateurs à mettre en place :

1. **Indicateurs concernant les personnes**
2. **Indicateurs concernant les process**
3. **Indicateurs concernant les technologies**
4. **Indicateurs concernant les données**

Nous présentons ci-après quelques exemples de tâches et d'indicateurs pour chaque catégorie.

- **Indicateurs concernant les personnes**

Recrutement	Formation	Participation
• Identification des ressources • Embarquement	• Plan de formation ressources • Développement du contenu des formation • Tableau d'avancement (% de personnes formées)	• Présentéisme • Création de livrables • Qualité des livrables
Indicateurs & KPI	# de décisions du Groupe de Travail sur la Data Gouvernance (GTDG) suivies par le comité de pilotage # de projets approuvés par le GTDG # de problèmes remontés au GTDG et résolus # de Data Owner identifiés # de Data Managers identifiés KPI # Taux d'adoption du programme de Data Gouvernance par le personnel (Enquête)	

- **Indicateurs concernant les process**

Création	Déploiement	Exécution
• Identification des process • % crées • % validés	• Création du plan de déploiement • % de process déployés • Suivi du planning et des coûts (VS plan)	• Compréhension et adoption • Usage • Efficience et efficacité • Backlog
Indicateurs & KPI	# de process de consolidation de données # nombre de normes, politiques et processus approuvés et mis en œuvre % nombre de définition de données et comment ces données sont utilisées dans les différents process # existence et adhésion à un processus d'escalade de demandes métier pour gérer les litiges relatifs aux données KPI # **Intégration d'indicateurs de data gouvernance dans les projets structurants**	

- **Indicateurs concernant les technologies**

Intégration	Efficacité	Catégories
• Identification des process • % crées • % validés	• Amélioration des données • Amélioration des process • Amélioration de la compréhension	• Data Qualité • Référentiel de données (MDM/ BIM…) • Metadonnées • Intégration de données • Architecture des données
Indicateurs & KPI	# de sources de données consolidées # de cibles data utilisant des données maîtrisées Intégrité des données à travers tout le SI # de data lineage documentés et compris # de Data Managers identifiés KPI **# présence et utilisation d'un identifiant unique**	

- **Indicateurs concernant les données**

Suivi et priorisation des entités de données communes Indicateurs de progrès	Suivi et priorisation des entités de données spécifiques Indicateurs de résultats	Indicateurs des process
• Réclamations sur les données issues de process documentés • Identification des réclamations non adressées • Suivi des temps de réponse	• Qualité des données. Les 7 dimensions de la qualité* • Efficience des process	• Taille du backlog • Mesure de la quantité de demandes/process en cours vs Backlog
Indicateurs & KPI	% d'améliorations effectuées % de réduction des duplication de données % exactitude des données Enregistrement/age des données hors cible KPI: **Augmentation de la productivité/ diminution des coûts** **Matrice de la qualité des données**	

Les 7 dimensions de la donnée

1. **Accessibilité** : logiciels et activités liés au stockage, à la récupération ou à l'action sur des données stockées.
   Surveiller et mesurer ; comment les responsables et les utilisateurs ont régi l'accès aux données dont ils ont besoin ont été en mesure de fournir un accès régi.
2. **Précision** : valeurs de données correctes et état du monde réel. Transformation d'origine, compréhension de la provenance et du linéage.
   Surveiller et mesurer : comment les données changent avec le temps et comparent aux normes de l'industrie ou de l'organisation.
3. **Complétude** : mesure dans laquelle les attributs de données attendus sont complets, à la demande du consommateur de données ou de l'organisation (les données peuvent être complètes, mais pas précises).
   Surveiller et mesurer : mesurez l'exhaustivité des données et des métadonnées. La dimension n'a pas besoin d'être complétée à 100% mais doit correspondre aux attentes et aux politiques.
4. **Cohérence** : les données correspondent-elles à des valeurs de données et des ensembles de données ? Gartner décrit la « cohérence des données entre les points de données proches ». Par exemple, si Chicago et Louisville ont des températures comprises entre 30° et 32°. Il est peu probable que la température à Indianapolis soit de 70°.
   Surveiller et mesurer : uniformité des métriques de qualité dans l'ensemble de l'organisation.
5. **Pertinence** : la proximité entre les besoins des consommateurs de données et le fournisseur de données qui permet d'utiliser des données avec une efficacité maximale.
   Surveiller et mesurer : données adaptées à l'usage prévu ; pourcentage de toutes les données requises divisé par toutes les données fournies
6. **Rapidité d'execution** : mesure dans laquelle les données sont suffisantes, à jour et disponibles en cas de besoin.
   Surveiller et mesurer : la disponibilité des données par rapport au temps requis par le consommateur.
7. **Validité** : conforme aux définitions définies et aux politiques / règles : une valeur peut être valide mais toujours inexacte
   Surveiller et mesurer : comparer les données et la politique (format, type, plage).

Enfin nous terminerons ce chapitre avec la démarche globale à adopter pour définir et mettre en place les « bons » indicateurs, le graphique ci-après illustre le process proposé :

Problèmes	Objectifs	Indicateurs et KPIs	Impacts
• Quels sont les problèmes à résoudre ? • Pourquoi cela pose problème ? • Qu'est-ce qui est important ? • Quels sont vos objectifs ?	• Quels changement voulez-vous voir ? • Comment cela va vous impacter ? • Qu'est-ce qui est important ? • Quels sont vos objectifs ?	• Quels process sont impactés • Quelles informations& données sont utilisées par ces process • Quelles améliorations nécessaires ?	• Bénéfices de l'amélioration des données pour atteindre les objectifs • Les changements positifs générés par l'acculturation

Cas d'applications

La gouvernance des données s'impose à toutes les organisations et à tous les niveaux de l'entreprise. Nous citons ci-après quelques exemples reposant sur des entreprises ayant déjà pris la mesure du changement à venir. Citons par exemples de secteur, de thématiques ou de métiers pour lesquels l'émergence récente des outils et des capacités pour reconnaître et traiter des données souvent autrefois inexploitables va profondément changer la donne en termes de création de valeur et d'organisation.

Relation Client

Exemple de données :

- Clients (noms, attributs,...)
- Données textuelles (Mails/Réclamations/réseaux sociaux/appels traduits en texte)
- Contrats

Cas d'usage :

- CRM (Analyse de « *churn* »/ rétention client...)
- Campagnes marketing

BTP/centres commerciaux/gare/aéroports

Exemple de données :

- Géolocalisation temps réel
- Données BIM
- Images
- IoT

Cas d'usage :

- Analyse et identification des flux de personnes
- Maintenance prédictive
- Surveillance d'infrastructure
- Consommation énergétique
- Web to shop...

Institutions Financières (Banques, assurances...)

Exemple de données :

- Finance (capital économique, ratios de solvabilité, comptabilité, actifs)
- Risques (provisions, market risk, criminalité financière)
- Clients (nom, attributs, ...)

Cas d'usage :

- Reportings automatisés Finance & Actifs, qualité de données
- CRM, KYC (Know Your Customer), analyse temps réel et prédictive des usages/tendances
- Campagnes marketing,
- Financial crime, Regulatory & Risk Compliance

Industrie, Juridique, transport, luxe...

Cas n°1 : mettre en place une gouvernance de la donnée en partant des exigences réglementaires autour de la qualité des données

Un grand groupe international du secteur banque assurance a sollicité Helis afin de l'accompagner dans la mise en œuvre d'un programme Data Gouvernance visant d'une part, à respecter les exigences réglementaires (Bâle III, Solvency II) en matière de qualité et traçabilité des données et d'autre part, à répondre aux besoins de gestion des données définis dans la politique du groupe.

Helis, accompagné par son partenaire Myriad, a élaboré une stratégie de gouvernance pour atteindre les objectifs précités qui s'appuie sur la mise en place d'un dispositif de maîtrise de la qualité des données.

Le renforcement des exigences bancaires après 2008 concernant la traçabilité et la qualité des données explique le point de départ des directions Financières & Risques dans la volonté de maîtriser le cycle de vie de la donnée. Dès lors qu'une donnée sensible apparaît dans un reporting de risques ou un bilan financier, la société doit être en mesure de démontrer la conformité ou l'absence de risques inhérents à la manipulation de cette donnée. En résumé, le respect des règlementations locales comme internationales a certes imposé mais donné également l'opportunité aux sociétés de casser les silos autour de la gestion de la donnée et d'en améliorer la gouvernance.

En outre, les opportunités sous-jacentes d'exploitation des données, concomitant au développement des infrastructures de stockage et de traitement de grands volumes de données, ont permis d'avancer des raisons supplémentaires en faveur de la mise au point d'un dispositif de maîtrise des données.

Ce dispositif repose sur la construction :

- d'une **organisation**, qui permette d'identifier les acteurs manipulant les données (Data Owners, Data Producers), qui définisse les rôles et les responsabilités de ces parties prenantes et des acteurs en charge de gérer le dispositif (Chief Data Officer et son équipe) à l'aide d'instance de pilotage (comitologie) et d'une charte de gouvernance.
- de **processus** qui permettent la mise en œuvre de la politique data management du groupe, à savoir identifier les processus métiers clés et IT à maîtriser, les données sensibles à cartographier, définir les règles de gestion de ces données, les processus d'évaluation de la qualité des données, les processus d'évolution de la plateforme de gestion des données.
- **d'outils**, qui peuvent être manuels dans un premier temps et qui assurent la mise en œuvre des processus de gestion de la qualité des données :
  1. des dictionnaires de données élémentaires et agrégées proposent une fiche d'identité de ces données (quel acteur produit quel contenu pour quel consommateur), un mode opératoire de fabrication, une étude de la criticité et du caractère confidentiel ;
  2. des data lineages qui cartographient les données utilisées et produites le long des processus métiers clés et IT, en relation avec les acteurs transverses (Risques, Actuariat, Finance, Marketing, Business Services, Comptabilité, DSI, Contrôle de gestion, Contrôle interne, etc) ;
  3. des indicateurs sont définis pour suivre l'avancement des processus cartographiés et la qualité des données sensibles inventoriées (Scorecards, Dashboards). Les résultats des requêtes testant la qualité de données élémentaires ainsi que les jugements d'expert des données sur la base des critères exactitude, complétude, appropriabilité et ponctualité ont permis de générer des tableaux de bord évaluant la qualité des données sensibles et des processus métiers clés.

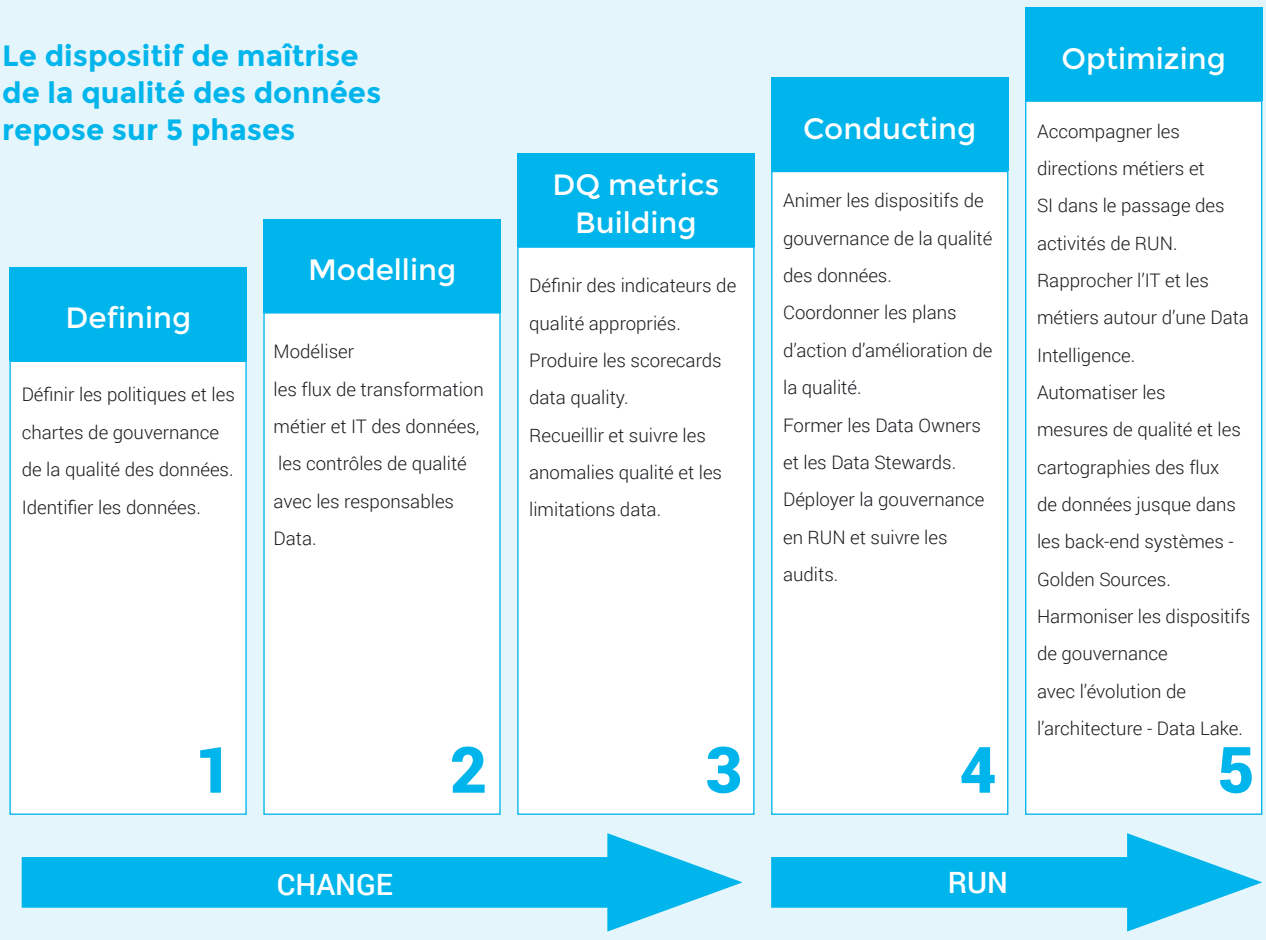
Une fois la phase de construction effectuée (phase CHANGE), Helis a accompagné la phase RUN du dispositif en veillant à la mise en place d'une démarche d'amélioration continue pour :

- **Aligner le fond et la forme** des instances de gouvernance (comités), l'utilisation des outils (dictionnaires, requêtes de test qualité, tableaux de bords de qualité) sur les exigences requises par la politique et les process data gouvernances mis à jour régulièrement par le groupe ;
- **Etendre le périmètre des données** Risque & Finance cartographiées et évaluées au périmètre des données Clients ;
- **Suivre les plans de correction des anomalies** et les plans d'évolution des limitations de données ;
- **Suivre les audits internes ou externes** sur la gouvernance des données ;
- **Suivre les chantiers d'automatisations des outils** et notamment d'évolution de l'architecture des données concernant :
  - le stockage et l'exploitation (nouvel environnement Hadoop incluant la création d'un lac de données – data lake – et l'apport des capacités d'analyse prédictive),
  - La cartographie de tous les flux de données y compris dans les systèmes IT (automatisation des data lineages),
  - L'optimisation du suivi de qualité et de mise à disposition des données (automatisation des tableaux de bords qualité et de visualisation).

En conclusion, des besoins clairement identifiés autour de la qualité des données ont permis de créer un cadre de gouvernance de la donnée. Celui-ci a ensuite rendu possible le suivi centralisé de nouveaux chantiers data complémentaires à celui du suivi de la qualité, en s'appuyant notamment sur des indicateurs de :

- **Personnes** (organisation définie),
- **Processus** mis en œuvre pour la qualité des données et à faire évoluer en fonction des nouveaux besoins de gestion des données, de perfectionnement des outils existants,
- **Données et Technologies**, qui qualifient le niveau de gouvernance des données et veillent à son amélioration continue.

Le dispositif de maîtrise de la qualité des données repose sur 5 phases



Cas n°2 : Mettre en place une gouvernance de la donnée afin de permettre à 5 marques de voiture d'utiliser une plateforme digitale commune pour leurs 270 sites internet

Un grand groupe international du secteur automobile a sollicité Helis afin de l'accompagner dans la mise en œuvre d'un programme de Data Gouvernance visant d'une part, à permettre aux différentes marques de connecter leurs systèmes d'informations sur un socle technologique commun et d'autre part, à répondre aux besoins de gestion des données définis dans la politique de l'alliance de marques.

Helis a élaboré une stratégie de gouvernance, qui s'appuie sur la mise en place d'un dispositif de maîtrise de la qualité des données, pour atteindre les objectifs précités.

Les constructeurs automobiles de l'alliance lancent un projet de transformation digitale qui représente la première collaboration entre les marques. Les enjeux sont à la fois politiques, chaque membre de l'alliance souhaitant garder la maîtrise de ses outils vis-à-vis des autres marques, mais également un chantier de transformation digitale jamais atteint en termes de cible et d'envergure puisqu'il couvre plus de 113 pays et 270 sites internet au total. La principale difficulté réside dans le fait que chaque marque possède sa propre infrastructure technologique. Ainsi la performance des systèmes varie énormément d'un constructeur à l'autre. L'objectif est donc d'avoir un socle technologique commun qui récupèrera la data, quelle que soit l'architecture de la marque, à l'aide d'un cache API capable d'épouser des formes de données différentes et de l'ingérer pour que la donnée de sortie soit la même pour le front client. La bonne mise en place de la gouvernance de la donnée est rapidement devenue le principal enjeu de ce projet, conditionnant sa réussite. En effet, même si la mise en place d'une infrastructure pour le stockage de grands volumes de données pour le socle commun technologique était également important, la nécessité d'un dispositif de data gouvernance pour la maîtrise de bout en bout de la donnée était critique.

Comme dans le cas d'usage décrit précédemment ce dispositif repose à la fois sur la construction :

- d'une **organisation**, qui permette d'identifier les acteurs, qui définisse les rôles et les responsabilités avec une instance de pilotage et une charte de gouvernance.
- de **processus** qui permettent la mise en œuvre de la politique data management, à savoir identifier les processus métiers clés et IT à maîtriser, les données sensibles à cartographier, définir les règles de gestion de ces données, les processus d'évaluation de la qualité des données, les processus d'évolution de la plateforme de gestion des données.
- **d'outils** qui assurent la mise en œuvre des processus de gestion de la qualité des données :
  - **Dictionnaires de données**
  - **Data lineages**
  - **Indicateurs**

Comme précédemment mentionné, une fois **la phase de construction effectuée** (phase CHANGE), Helis a accompagné la phase RUN du dispositif en veillant à la mise en place d'une démarche d'amélioration continue pour :

- **Aligner le fond et la forme** des instances de gouvernance (comités) ;
- **Étendre** le périmètre des données Client cartographiées et évaluées a**u périmètre des données Marketing** ;
- Suivre les plans de correction des anomalies et les plans d'évolution des limitations de données ;
- Suivre les audits internes ou externes sur la gouvernance des données ;
- Suivre les chantiers d'automatisations des outils et notamment d'évolution de l'architecture des données concernant :
    - Le stockage et l'exploitation des données,
    - La cartographie de tous les flux de données y compris dans les systèmes IT,
    - L'optimisation du suivi de qualité et de mise à disposition des données.

En conclusion, des besoins clairement identifiés autour de la qualité des données ont permis de créer un cadre de gouvernance efficace pour la bonne gestion de la donnée. Celui-ci a ensuite rendu possible le suivi centralisé de l'utilisation de la data, des actions reliées à celle-ci et de sa qualité en s'appuyant notamment sur des indicateurs de :

- Personnes (organisation définie),
- Process, mis en œuvre pour la « data quality » et à faire évoluer en fonction des nouveaux besoins de gestion des données et de perfectionnement des outils existant,
- Données et Technologies, qui qualifient le niveau de gouvernance des données et veillent à son amélioration continue.

Cas n°3 : Normalisation de la donnée digitale/ comportementale pour la mise en place du projet Client 360°

Un grand groupe international du secteur agro-alimentaire a requis les services de nos consultants afin de l'accompagner dans la mise en œuvre de son projet de centralisation des données clients visant à répondre aux besoins de gestion des données définis dans la politique du groupe.

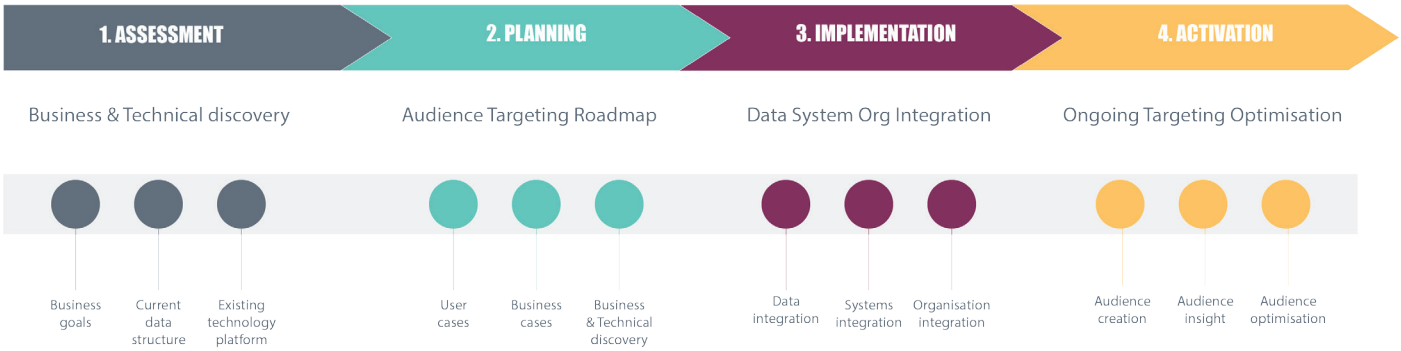
Helis, a élaboré une stratégie de gouvernance pour **la normalisation de la donnée digitale comportementale** pour les 13 marques du groupe. L' objectif étant une gestion de la donnée centralisée pour l'ensemble des marques et des entités du groupe et une facilité de traitement pour agréger la donnée au sein des profils clients omnicanal (DMP).

Le groupe a entamé des travaux de centralisation de la donnée client dans l'objectif de réussir à déployer en parallèle un outil de CRM (Sales Force) et une DMP. De multiples acquisitions de marques au cours des dernières années ayant créé une forte disparité des typologies de données, nous avons immédiatement identifié un risque très fort sur la mise en place du projet dans les délais impartis.

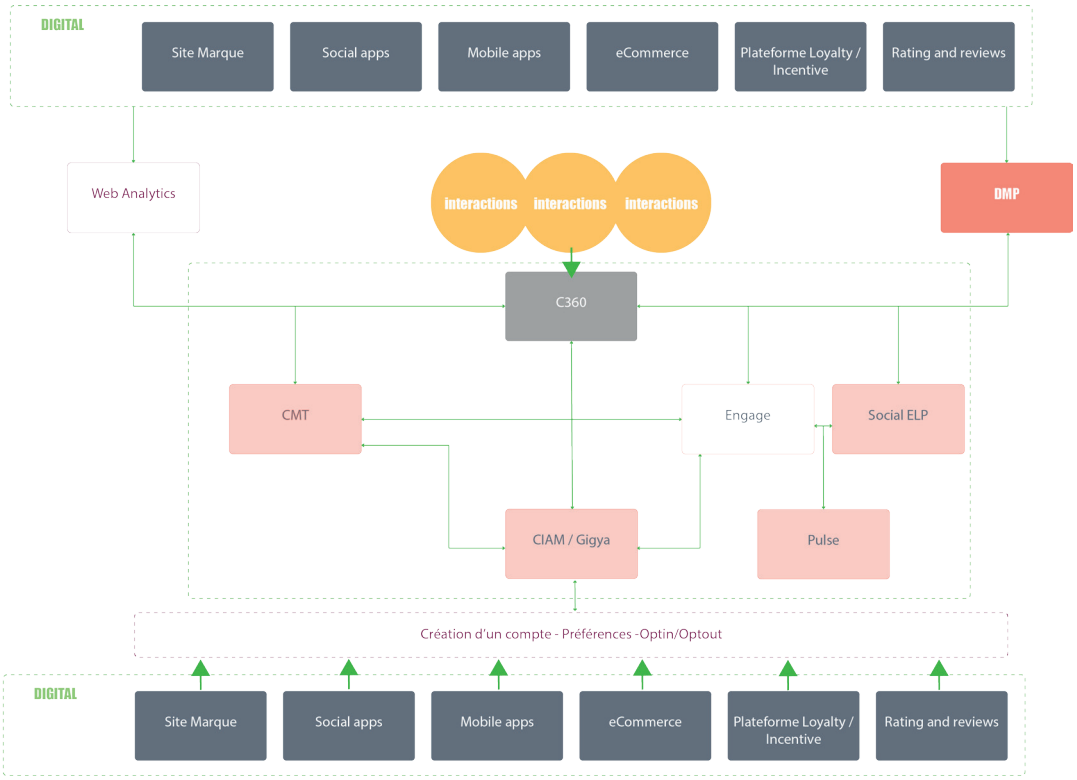
Les 13 marques du groupe avaient chacune leur communication omnicanale avec leur propres bases prospects et clients. De multiples actions marketing étaient menées en silos avec plusieurs intermédiaires. La gestion web analytics était décentralisée, et ses administrateurs, à la fois externes et internes, étaient sans contrôles sur les rôles et responsabilités. En outre, les propriétaires des données des marques étaient pour la plupart, non identifiés. De ce fait, la capacité du client à réussir la mise en œuvre d'un déploiement de CRM et d'une DMP centralisée dépendait directement de sa capacité à piloter la mise en œuvre d'une gouvernance pour la normalisation et la maîtrise de bout en bout de la donnée des 13 marques.

Comme pour les cas d'usage précédents, notre démarche nous aura permis :

- D'identifier les différents **rôles et responsabilités des utilisateurs** sur les outils de gestion et manipulation de la donnée digitale.
- De normaliser **la nomenclature de nommage** des champs de données, de l'utilisation standard de variables personnalisées et d'une configuration unique standardisée pour l'implantation sur les sites des marques.
- D'identifier les **processus** à mettre en place pour la gestion des demandes de nouvelles fonctionnalités ou de changements de l'existant.
- D'identifier les **personnes ressource** pour chacune des marques.
- D'identifier les **outils** à utiliser pour la préparation de la donnée et son transfert aux bases de données de grand volume.
  - **Dictionnaires de données**
  - **Data lineages**
  - Des **indicateurs** d'évaluation des performances marketing des sites web
  - De **visualisation de la donnée** pour les rapports et tableaux de bord.



Champs d'intervention



Cartographie des flux de données

Également comme les deux précédents cas d'usage, une fois la phase de construction effectuée (phase CHANGE), Helis a accompagné la phase RUN du dispositif en veillant à la mise en place d'une démarche d'amélioration continue pour :

- Aligner le fond et la forme des instances de gouvernance (comités) ;
- Étendre le périmètre des données comportementales Client cartographiées et évaluées au périmètre des données Marketing et de profils clients multi-sources « client 360° » ;
- Suivre les plans de correction des anomalies et les plans d'évolution des limitations de données ;
- Suivre les audits internes ou externes sur la gouvernance des données ;
- Suivre les chantiers d'automatisations des outils et notamment d'évolution de l'architecture des données concernant :
  - Le stockage et l'exploitation des données,
  - La cartographie de tous les flux de données y compris dans les systèmes IT,
  - L'optimisation du suivi de qualité et de mise à disposition des données.

En conclusion, le besoin de normalisation de la donnée comportementale aura permis de mettre en œuvre et de réussir un chantier de gouvernance de la donnée appliqué au projet « client 360° ». Ce cas d'usage, d'un périmètre limité, aura permis aux équipes de projets de préparer une gouvernance de la donnée pour la centralisation de la donnée client sur un stockage de grande capacité. Ce cadre de gouvernance a cependant rendu possible le suivi centralisé de nouveaux chantiers data complémentaires à celui du suivi de la donnée comportementale, en s'appuyant notamment sur des indicateurs de :

- Personnes (organisation définie),
- Process, mis en œuvre pour la « data quality » et à faire évoluer en fonction des nouveaux besoins de gestion des données, de perfectionnement des outils existant,
- Données et Technologies, qui qualifient le niveau de gouvernance des données et veillent à son amélioration continue.

Remerciements

L'auteur remercie pour leurs apports et contributions, sans qui ce livre blanc n'aurait pu voir le jour :

- Pierre Samsom, Consultant Senior Data HELIS qui a mis en œuvre chez nos clients les cas d'usage 2 et 3
- Xavier Ros, Chef de Projet Data Gouvernance MYRIAD
- Thibault Bourdel, Responsable Développement chez BIMTech

À propos

helis

À propos de Helis

Créée en 2004, **Helis est un cabinet de conseil en AMOA** qui intervient auprès de clients Grands Comptes dans le domaine de l'IT, des télécoms et la donnée.

Helis s'est imposée comme un partenaire de confiance grâce un engagement fort : apporter spécifiquement à chaque client ou projet une valeur ajoutée maximale et maîtrisée, en recrutant « sur-mesure » des profils expérimentés dotés d'une double compétence, technique et métier.

Dans le domaine « data », Helis répond plus particulièrement aux enjeux des directions métiers, fonctionnelles et SI en replaçant la donnée au cœur de leurs projets et en facilitant leur collaboration. En actionnant la donnée, Helis accompagne ses clients dans l'identification des leviers qui transforment l'entreprise vers le succès.

À propos de l'auteur

Diplômé du MBA de l'ISG, **Charles-Eric de La Chapelle**, est un entrepreneur qui a initié ou contribué au développement d'entreprises en France et à l'étranger dans l'informatique, les telecom, l'internet, la formation, le Big Data et l'Intelligence artificielle (IA). Passionné par les nouvelles technologies, il est président et fondateur de Myriad, french tech spécialisée en Intelligence Artificielle fondée en 2016, dont il dirige l'activité de conseil et les programmes de R&D.

Depuis 2016, Helis et Myriad ont établi un partenariat étroit portant sur les activités de conseil autour de la donnée.

« Au-delà des infrastructures et de la qualité et de la pertinence des modèles algorithmiques, Je suis convaincu que le succès des projets Big Data et IA dépendent en premier lieu de la transversalité de l'approche dans les organisations et de l'implication et de la responsabilisation des hommes et des femmes à tous les niveaux pour faire évoluer les mentalités et les processus »