



Livre Blanc

Feuille de route d'une stratégie de Data Management

Facteurs de succès – Bonnes pratiques – Technologies





Abdelaziz Joudar,
Président co-fondateur
de DataValue Consulting

Edito

Feuille de route d'une stratégie de Data Management

Dans le climat actuel de crise et d'incertitude, la mise en place d'une stratégie data, combinée avec des technologies d'Intelligence Artificielle, devient vitale pour les entreprises en quête de croissance et de compétitivité.

Beaucoup d'acteurs investissent depuis plusieurs années dans des programmes data, mettant en place des cas d'usage et des socles technologiques complexes. Ces investissements ne sont pas sans risques. D'autres entreprises, conscientes de l'importance de ce virage technologique, se lancent depuis peu dans l'élaboration d'une stratégie de Data Management.

Cette série de livres blancs a vocation à vous guider à chaque étape de votre projet de Data Management. Nous vous livrons nos convictions, avec pragmatisme et audace, sur les actions à mener pour réussir votre projet data. Cet ouvrage est le résultat d'une analyse et d'une restitution des expériences acquises par nos consultants dans le cadre des projets data que nous avons mis en place dans plus de 100 entreprises.

Au-delà de l'expertise, ces convictions de Datavalue Consulting démontrent notre volonté d'aider nos clients sur des thématiques d'actualité. Si ces convictions vous font réagir, parlons-en... ensemble.

Sommaire

PAGE 5

➤ 1. Bâtir sa stratégie data

PAGE 5 – 6

Enjeux, cas d'usage et méthodologie pour construire une stratégie data-centric.

PAGE 7

➤ 2. Construire sa plateforme de Data Management

PAGE 7 – 11

Mise en place d'une gouvernance des données (rôle, processus, organisation).

PAGE 12 – 14

Démarche de gestion des référentiels de données.

PAGE 15 – 19

Processus d'amélioration de la qualité des données.

PAGE 20 – 21

Stratégie de tagging des données.

PAGE 22 – 23

Anonymisation et protection des données.

PAGE 24 – 25

Stratégie de partage des données avec des fournisseurs tiers.

PAGE 26 – 27

Modèles d'organisation des données.

PAGE 28

➤ 3. Pour aller plus loin

PAGE 28 – 30

Glossaire et définition des mots-clés.

PAGE 31

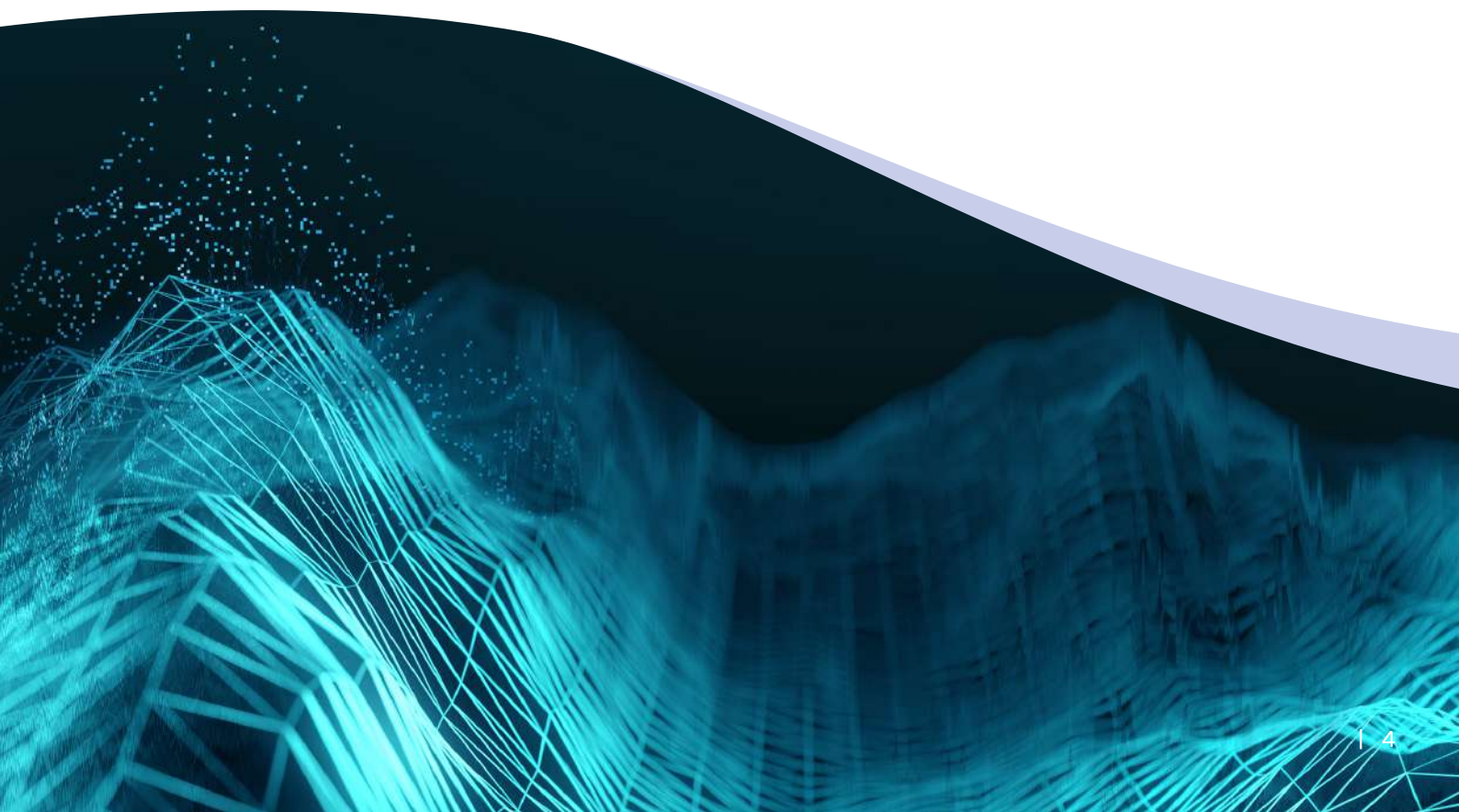
Pour aller plus loin.

Introduction

Les solutions résident en nous.

La numérisation exponentielle de notre économie, les avancées technologiques majeures récentes et les possibilités de valorisation offertes par les outils d'analyse et d'Intelligence Artificielle sont des facteurs qui renforcent l'importance de la data, considérée comme l'un des actifs les plus importants d'une organisation. Il devient urgent pour les entreprises d'apprendre à maîtriser et exploiter ce capital data.

Beaucoup d'entreprises choisissent d'investir massivement dans des solutions technologiques. Mais cet acte d'acquisition, aussi séduisant et satisfaisant soit-il, n'a jamais résolu les problèmes de fond. Notre conviction est la suivante : les entreprises doivent cesser de réfléchir « technologie » et avant tout réfléchir aux cas d'usage afin de développer des « capacités » mêlant pratiques, processus et organisation.

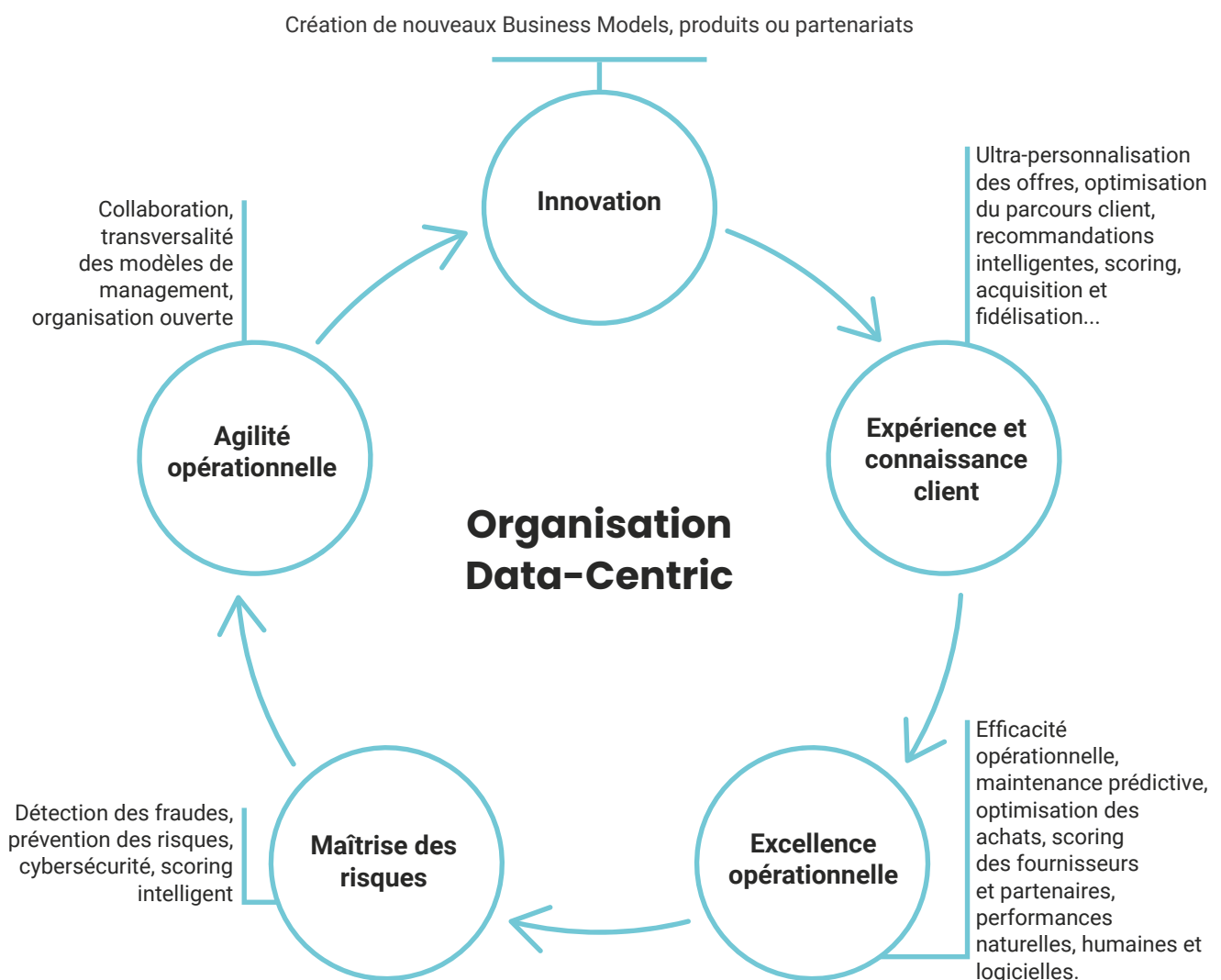


1 Bâtir sa stratégie data

Si les entreprises demeurent de nos jours majoritairement « Process-Centric » en définissant des processus métiers qui s'exécutent de bout en bout quelles que soient les données manipulées, on constate une accélération des transformations vers des entreprises « Data-Centric » centrant ces dernières sur leurs clients et collaborateurs malgré le volet technologique d'une telle opération.

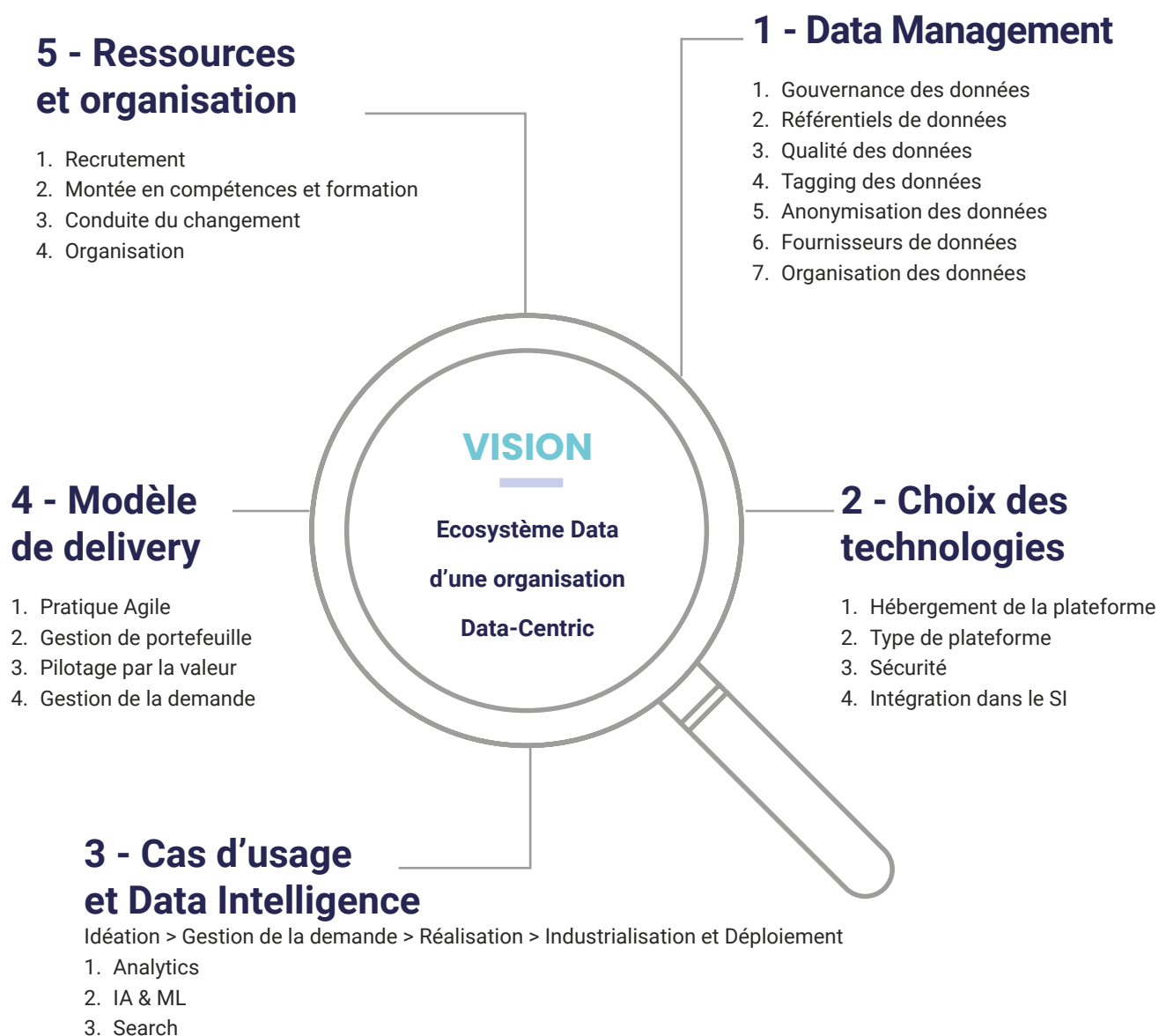
Ces stratégies « Data-Centric » mettent les données au cœur de la création de valeur. Il ne s'agit pas uniquement de collecter et stocker ces données mais également de les transformer et les synthétiser pour expliquer les événements passés, définir des axes d'amélioration, des plans d'actions ou réaliser des prédictions. L'organisation « Data-Centric » ouvre donc un éventail illimité d'opportunités.

Avantages d'une organisation Data-Centric



Comment bâtir une organisation Data-Centric ?

Il s'agit d'initier une transformation globale avec **plusieurs volets** à consolider (cf. dessous).
Cette transformation globale démarre par la construction d'une vision de la transformation installant une culture, une ambition, un périmètre et des offres de services.



Si les entreprises demeurent de nos jours majoritairement « Process-Centric », on constate une accélération des transformations vers des entreprises « Data-Centric ».

2 | Construire sa plateforme de Data Management

Gouvernance des données

La gouvernance des données est un ensemble de pratiques et de processus qui contribuent à assurer la maîtrise du patrimoine de données au travers d'une organisation. Elle doit être pilotée par le métier et non l'IT et doit être le point de départ de la stratégie data.

Pour structurer un framework de Data Gouvernance au sein d'une organisation, nous considérons qu'il faut tout d'abord évaluer l'état de l'art sur les points suivants :

- La connaissance de son patrimoine de données,
- L'acculturation des différents acteurs, nécessaire à la mise en place d'une stratégie Data-Driven,
- La structuration des responsabilités et de l'ownership autour de la donnée,
- La mise en place de normes de qualité et la gestion quotidienne de celle-ci,
- La sécurisation des données les plus sensibles de l'entreprise,
- La maîtrise du cycle de vie de ses données,
- La diffusion de la donnée dans une logique de désilotage,
- La valorisation et la création de valeur autour de la donnée.

Lorsqu'une organisation se trouve au niveau 0 de la gouvernance des données, nous conseillons généralement de procéder par périmètre bien définis en travaillant sur des cas d'usage. Cette approche permet de mesurer rapidement les résultats du framework mis en place.

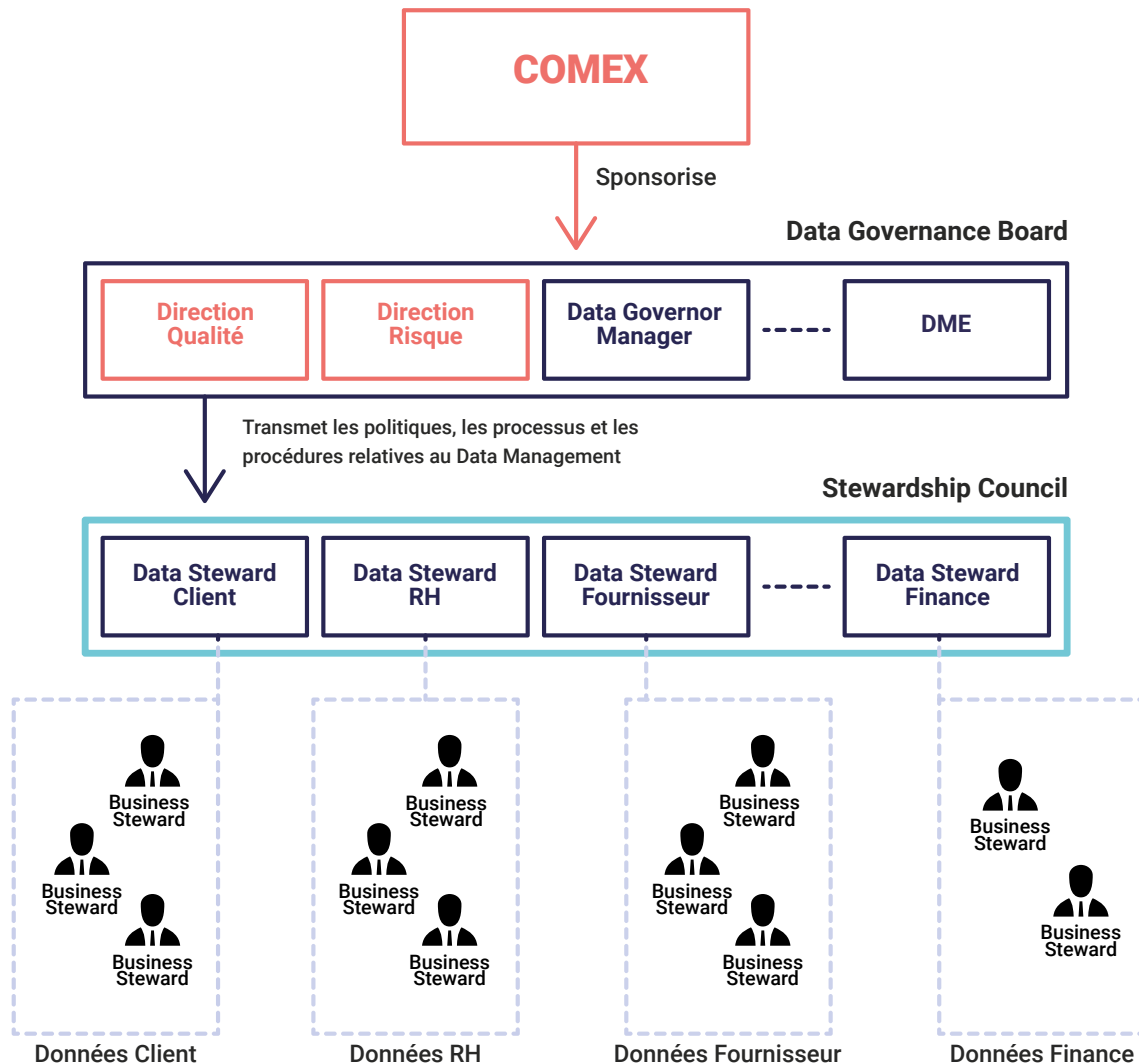
Les cas d'usage peuvent porter sur différents chantiers liés à la data :

- L'exploitation et la valorisation des données,
- Des initiatives permettant une acculturation transverse des acteurs de l'organisation,
- L'optimisation ou la définition de nouveaux processus,
- L'architecture et la gouvernance dans la mise en place de nouveaux applicatifs.

Une étape fondamentale à la mise en place d'une gouvernance est la définition de rôles et responsabilités autour de la donnée. ».

Chief Data Officer (CDO), Data Management Executive (DME) et Data Stewards (DS) sont autant de rôles permettant le management et le monitoring quotidien de son patrimoine de données. Ces derniers jouent un rôle essentiel dans la sensibilisation des acteurs autour des enjeux data. La création d'un modèle d'objets métier, de familles et de sous-familles de données est également indispensable afin de **réconcilier la vision de l'IT et du métier**.

Modèle de Data Gouvernance non-invasif par famille de données



Exécutif

- Autorise les changements de l'organisation
- Sponsorise les nouvelles procédures et le programme de changement

Tactique

- Propose des standards partagés
- Veille aux respects des standards et de la conformité
- Accorde des exceptions aux standards

Stratégique

- Définit les règles et procédures
- Identifie, nomme les stewards et approuve leurs actions
- Priorise les actions data

Opérationnel

- Gère les données au quotidien
- Définit les données et leurs usages pour leur Business Unit
- Nettoie les données

L'acculturation de l'ensemble des acteurs est un autre élément essentiel au succès d'un projet de gouvernance.

Une vision stratégique data partagée par tous garantit que les parties prenantes comprennent et assimilent les objectifs et la vision, acquièrent les compétences nécessaires et s'impliquent dans cette transformation.



Les outils collaboratifs de gouvernance des données (Collibra, Data Galaxy, AB Initio...), conçus pour être accessibles par le métier, sont un bon moyen de monitorer efficacement la donnée tout en incitant les différents acteurs à s'investir sur celle-ci.

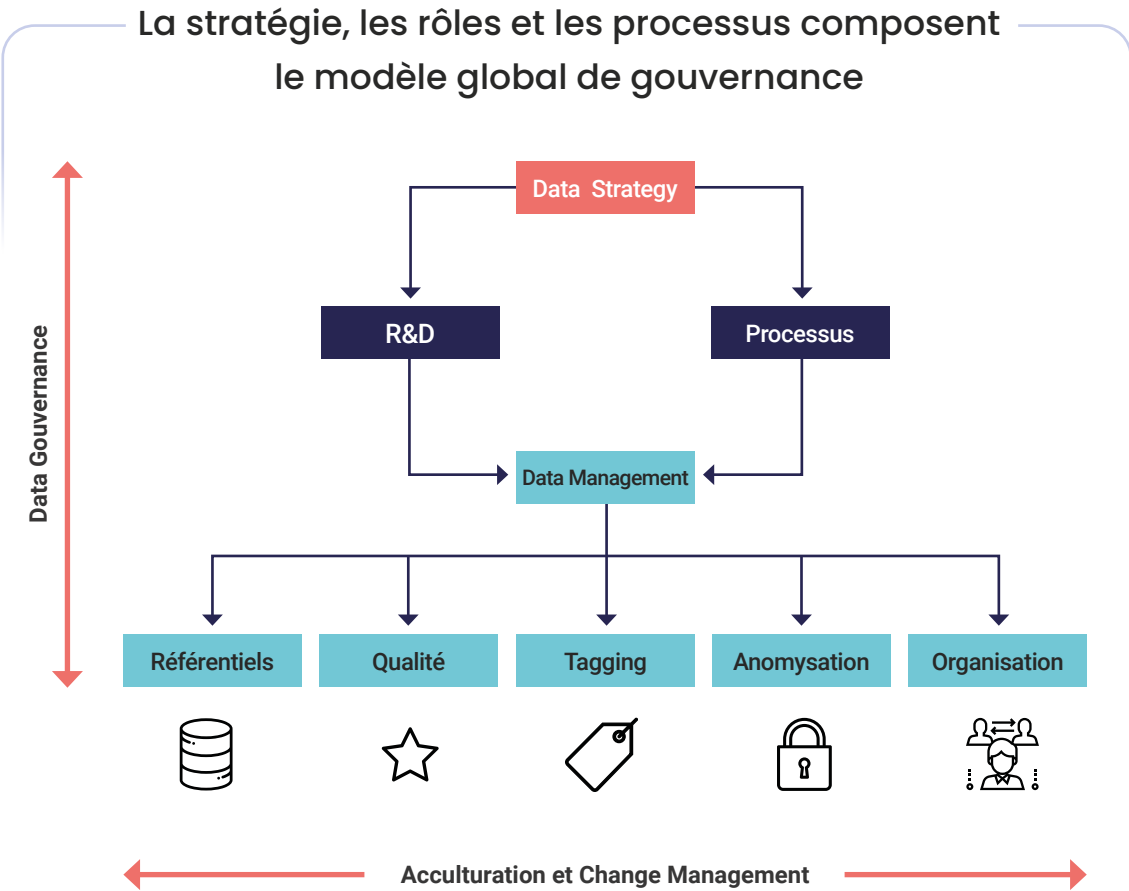
Les portails de type Data Lab centralisent de multiples jeux de données, et permettent d'explorer des cas d'usage de manière collaborative. Ils font émerger des communautés data axées sur différents sujets.



Des processus doivent également être instaurés afin d’assurer une traçabilité transverse au sein de l’organisation tout en facilitant **l’accès, l’utilisation et le partage de la donnée**. Le Data Officer, piloté par le CDO, est garant de la définition et de l’application de ces derniers.

Cet ensemble de rôles, responsabilités et processus permettront d’assurer un management efficace et transparent de la donnée à plusieurs niveaux :

- La gestion des référentiels,
- Le monitoring de la qualité des données,
- Le tagging des données,
- L’anonymisation des données
- L’organisation des données.



Nos convictions :

- L'acculturation de l'ensemble des acteurs est indispensable à l'amorce d'une dynamique de transformation par la data.
- L'enrichissement et la valorisation de la donnée doivent être menés de manière itérative.
- La gouvernance de la donnée doit être adaptée aux besoins métiers et aux enjeux de l'industrie concernée.
- La qualité de la donnée, sa diffusion et ses usages sont constitutifs de sa valeur et doivent se gérer proactivement et durablement.
- La gouvernance des données doit être pilotée par des experts et exécutée au quotidien par tout le monde.
- La gouvernance des données doit faire ses preuves. Nous recommandons une démarche par cas d'usage pour lancer cette transformation.

Nous recommandons une démarche par cas d'usage pour lancer une transformation vers une organisation Data-Centric.

Référentiels de données

Dans tout système d'information urbanisé et performant, la **brique « référentiels de données » (ou Master Data Management)** représente le pivot central qui harmonise la donnée à l'échelle de l'entreprise et au-delà. C'est aussi la pierre angulaire de la transversalité. C'est une brique qui gère le **cycle de vie des données de référence** et leur consommation à partir du point qui garantit leur véracité.

Cette brique doit permettre de :

- Cartographier les données de référence ainsi que la démarche de gestion et de réconciliation des données,
- Mettre en place une architecture adaptée aux besoins et contraintes des différents métiers de l'entreprise ainsi qu'à la nature des référentiels,
- Embarquer les métiers à la fois en matière de propriété et de gestion,
- Dérouler une approche à long terme même après la mise en production des premières itérations du projet.

NOS CONVICTIONS

- La **gestion des référentiels** dans le cadre d'une démarche MDM est une brique importante d'un socle data.
- C'est une **démarche transverse** à l'entreprise qui en plus de l'adoption de l'outil devra être alignée sur le contexte et les enjeux de l'entreprise. Elle nécessite un sponsorship fort et une vision transverse.
- Les outils MDM facilitent l'usage de la data. ils améliorent la **qualité des informations** délivrées aux autres systèmes (centralisation et mutualisation de l'information).
- La démarche de construction est **itérative** (par référentiel, par métier, par fonction...) et fondée sur le prototypage en partant d'un lot initial.
- Chaque référentiel (client, produit, organisation...) dispose de sa **roadmap propre** tout en assurant une cohérence globale.
- Le **développement de son propre MDM « maison »** paraît peu envisageable car les fonctionnalités à implémenter sont trop nombreuses et compliquées à maintenir.

Bonnes pratiques d'un projet de référentiel de données

● **Souhaitables** ● **Obligatoires**

Intégration

Mécanisme d'intégration des données :
Par lot, au fil de l'eau, scanning

Mapping :

Transformation, transcodification, gestion des erreurs,
Gestion des identifiants uniques, des références croisées

Exposition

Vue unifiée :

GUI, Web services, Requêtage,...

Mécanisme d'export des données :

Total, partiel, propagation...

Cycle de vie

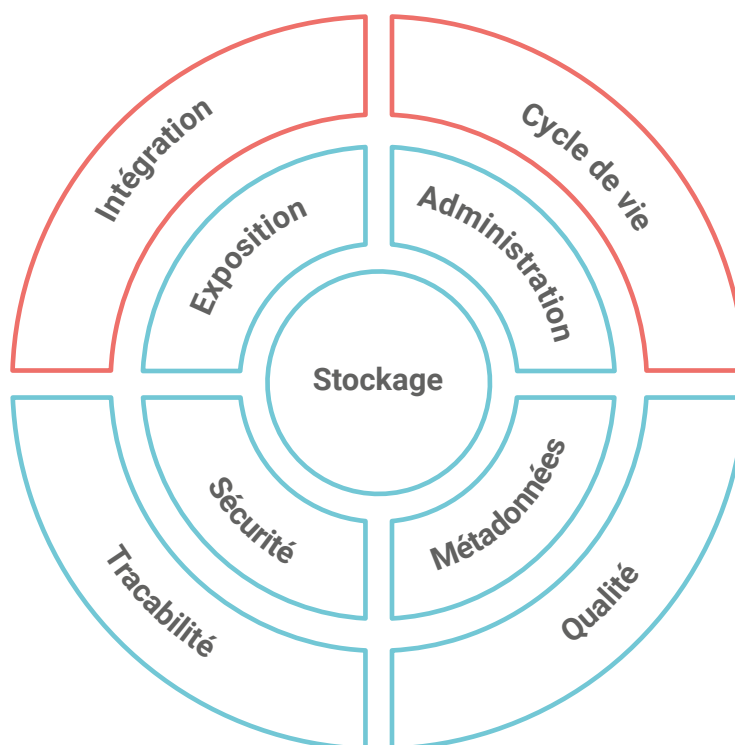
Cycle de vie / Durée de vie des données,
Marquage des données obsolètes / supprimées,
Dernière mise à jour / utilisation,
Conservation légale

Administration

Support de types de données évolués :
Données métier, composites, complexes

Outils de modélisation des données :
Versionning, documentation, rétro-ingénierie

Dictionnaire de données



Sécurité

Rôles et droits d'accès,
Règles de contrôle des accès,
Intégrité des données et archivage,
Anonymisation, pseudonymisation

Stockage

SGBD interne ou externe,
Type(s) de SGBD interne / supportés

Métadonnées

Multi-entités, multi-domaines,
Golden Sources et Golden Data,
Stewardship,
Workflows et BPM

Traçabilité

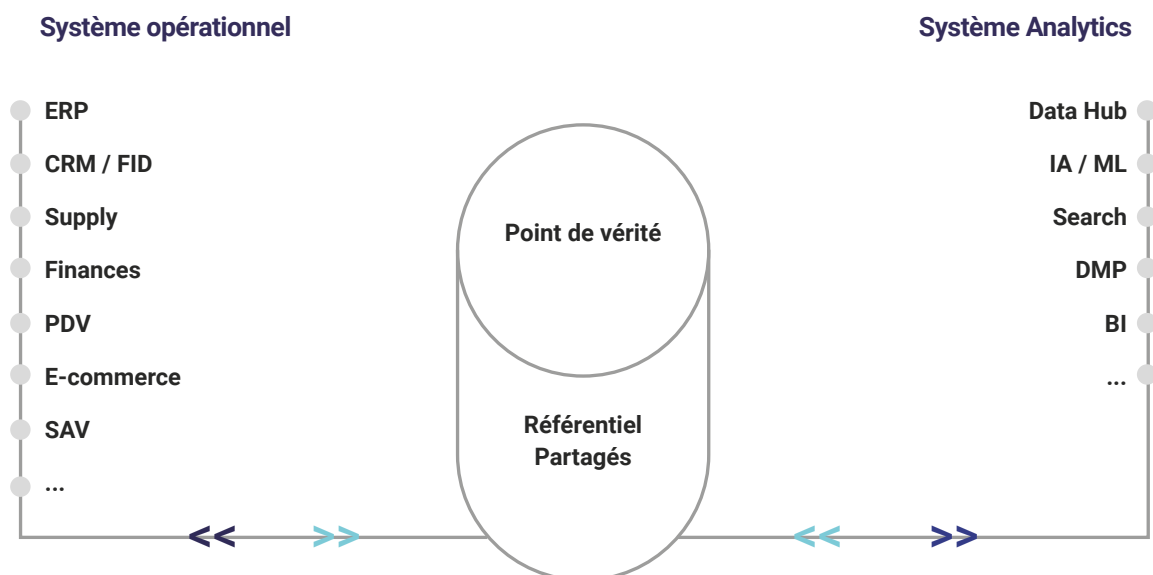
Data Linéage,
Historisation des données,

Qualité

Règles de conformité,
Doublons / Valeurs aberrantes,
Marquage et mise en quarantaine,
Analyse en continu,
Mise en quarantaine

Nos convictions :

- L'étude des solutions sectorielles pour le déploiement de solutions packagées avec un fort ROI est recommandée.
- Lors de la mise en œuvre du MDM, pensez « **bi-directionnel** », « **collaboratif** », « **évènementiel** », « **temps réel** », « **non structuré** ».



- Pour répondre aux nombreuses contraintes que connaissent les entreprises, la modélisation et la gestion multi-domaines est à promouvoir en lieu et place d'une spécialisation par domaine (Client, Produit, etc).
- Le déploiement de la gestion des référentiels sur le cloud est à recommander pour installer un modèle de service à fort engagement.
- L'intégration des référentiels à un socle Big Data crée un pivot réconciliant les données externes (souvent des données de paramétrage) et les données internes déjà harmonisées grâce au MDM.
- Le MDM développe aussi la forte disponibilité des données de l'entreprise pour mieux appréhender les stratégies multicanales.

Qualité des données

La **qualité des données** désigne les caractéristiques intrinsèques des données qu'elles soient internes ou externes à l'entreprise. Nous pouvons citer l'exactitude, l'exhaustivité, la cohérence, la validité, l'actualité, l'intégrité, la clarté et la sécurité, etc. Selon une enquête internationale exclusive menée par Experian Marketing Services, 92 % des entreprises françaises n'ont pas confiance dans la qualité de leurs données et que 27 % des données sont considérées comme erronées.

La qualité des données peut être dégradée à deux niveaux :

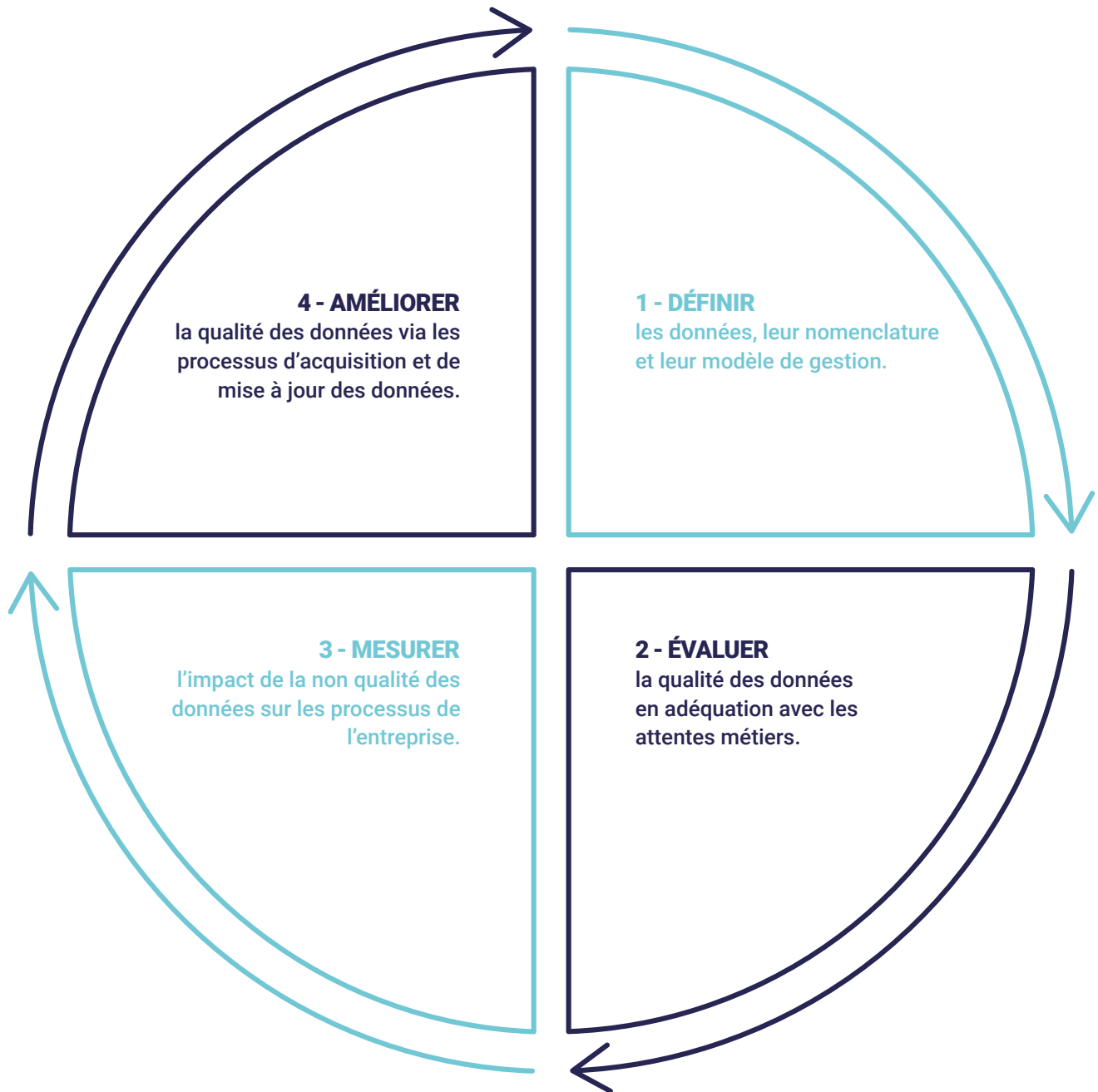
- La description des données : conflits entre noms d'objets ou imprécisions sur la définition des objets.
- La donnée en elle-même : valeurs nulles, doublons, valeurs anormales, données obsolètes.

La majorité des entreprises a déjà essayé de mettre en place un processus de gestion des données correspondant à la mise en place des bonnes pratiques et des techniques de mesure et de contrôle en qualité des données. Mais ces processus de gestion de données restent souvent cloisonnés, respectant les silos d'organisation existants et ne sont pas transverses au fonctionnement des entreprises.

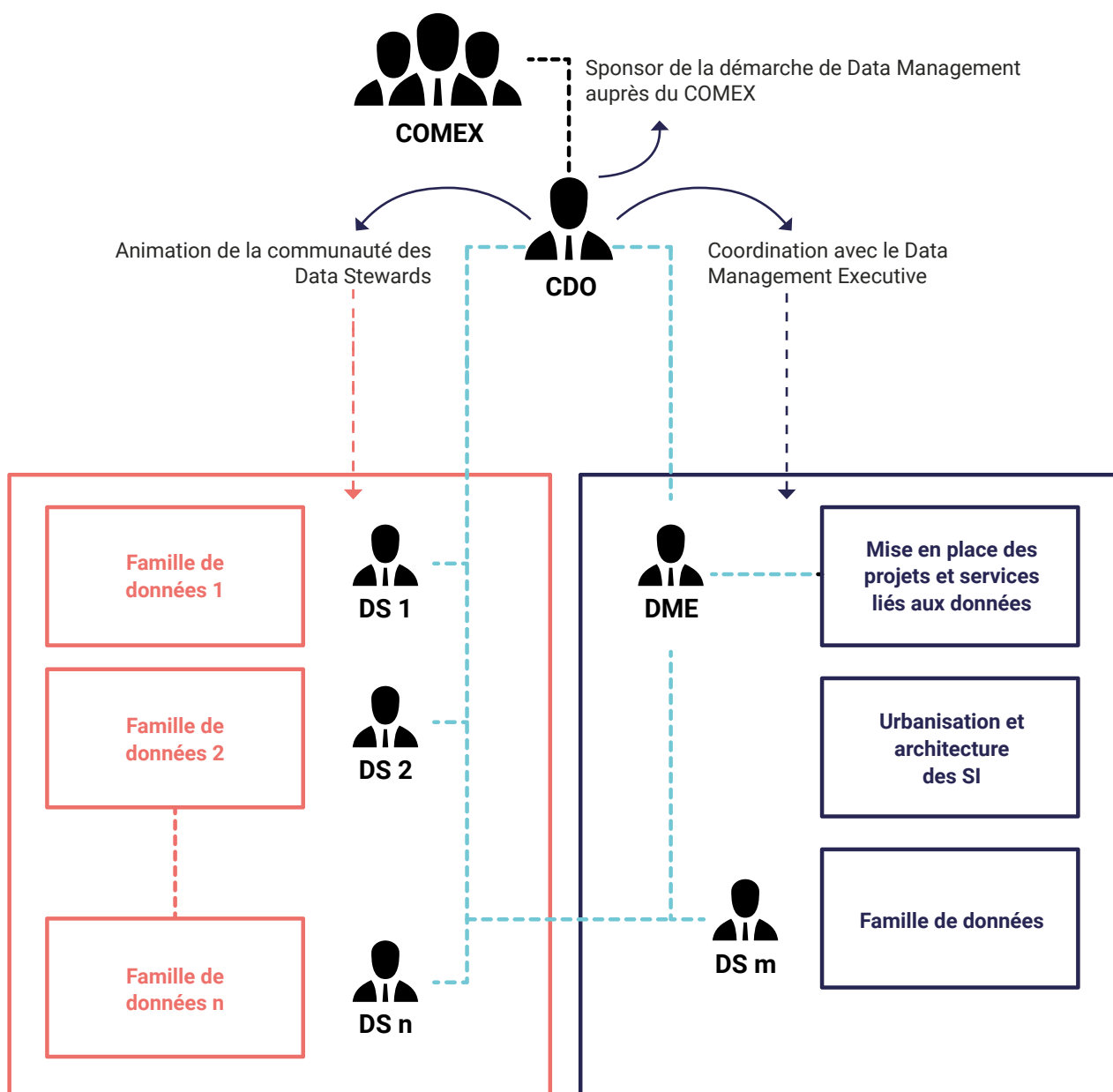
Nos convictions :

- Seule une **démarche organisationnelle** peut améliorer la qualité des données en adaptant les investissements aux enjeux de qualité.
- Cette démarche devra reposer sur **4 piliers** :
 - la stratégie,
 - les processus,
 - les collaborateurs,
 - la supervision / l'amélioration continue.
- Le processus de gestion de la qualité des données devra être itératif en s'appuyant sur **une mise à niveau de la qualité des données « by design »** vers un niveau de qualité standard défini dans la stratégie data. Il faudra donc définir des processus de prévention de la non-qualité en amont et des processus de remédiation des anomalies en aval.

Les étapes de gestion de la qualité des données



3 rôles clés sont à définir pour garantir la mise en œuvre de la qualité, conformément aux enjeux de l'entreprise et aux besoins d'exploitation par les métiers.



Description des 3 rôles clés

Chief Data Officer (CDO)

Il définit les règles de gouvernance, suit la mise en œuvre et coordonne l'ensemble des activités de Data Management.

- Développer et maintenir une stratégie data,
- Organiser la gouvernance des données,
- Définir et promouvoir le modèle de données d'entreprise,
- Sponsoriser les projets et les services de gestion des données,
- Communiquer et promouvoir les politiques, les standards et les procédures de gestion des données,
- Superviser et coordonner les activités de gestion des données,
- Suivre les coûts liés à la gestion des données et proposer des optimisations.

Data Management Executive (DME)

Il organise les services de gestion des données, en cherchant à développer le partage et la ré-utilisation pour plus d'efficacité et de cohérence.

- Définir et mettre en œuvre des architectures de données répondant aux besoins métiers,
- Définir et maintenir un cadre de mutualisation des services transverses liés aux données,
- Coordonner et piloter les projets d'évolution de l'architecture des données,
- Suivre les coûts liés à la gestion des données et proposer des optimisations,
- Garantir la mise en œuvre dans le SI des contrôles permettant le respect des exigences de qualité et des contraintes réglementaires.

Data Steward (par famille de données)

Il décrit les besoins de gestion des données dont il a la charge en fonction des usages, suit la mise en œuvre, et veille à délivrer des données de juste qualité.

- Comprendre les processus liés à la collecte de la donnée et à sa consommation,
- Recenser les besoins et les exigences relatifs aux données, les hiérarchiser, et proposer des solutions /services en coordination avec le DME,
- Mettre en œuvre les bonnes pratiques de gestion des données en coordination avec le CDO,
- Gérer et résoudre les problèmes liés aux données,
- Suivre les coûts liés au Data Management sur son périmètre et proposer des optimisations.



Nos convictions :

- **Les règles de mise en qualité** sont propres à chaque métier mais doivent obéir aux grands principes de la stratégie data et aux processus qui pérennisent les travaux de mise en qualité.
- **Les outils ne sont pas une fin en soi** (les référentiels de contrôle et l'évaluation de la qualité seulement le sont).
- Le développement de sa solution « maison » de gestion de la qualité de données ne semble pas pertinent car les fonctionnalités à implémenter sont nombreuses et compliquées à maintenir. De nombreuses technologies émergentes (qui ne sont pas encore arrivées à maturité) répondent aux problématiques de la gestion de la qualité des données.
- Le marché est dominé à 90% par 4 familles d'outils selon le rapport d'étude MarketsandMarkets™ : **Talend, Informatica, Collibra, Ab Initio**.
Le choix d'un outil nécessite au préalable de cartographier les usages et fonctionnalités à couvrir par la solution de gestion de la qualité des données, avant de lister les critères de choix et de les évaluer : richesse et pertinence des fonctionnalités, prix, pérennité de la solution, expérience utilisateur, intégration dans le SI, formation, reporting qualité... Les outils basés sur l'IA ne sont pas encore matures.
- Le maintien de la qualité des données nécessite **la mise en œuvre de bons KPI** pour mesurer :
 - Des critères intrinsèques aux données elles-mêmes,
 - Des critères de services liés à l'utilisation de ces données,
 - Des critères de sécurité liés à l'ensemble du dispositif de gestion des données.
- Au-delà des modèles, **la réalisation de documentations, de politiques, de standards** et de processus est primordiale pour la réussite d'un projet de gestion de la qualité des données.
- Au-delà des processus et des outils, **la montée en compétence des équipes** doit être prise en compte :
 - le mode agile sur cette brique data n'est pas aisé,
 - le modèle standard client/fournisseur n'est pas adapté à cause de la dimension temps réel des opérations,
 - La confiance entre les équipes est primordiale.

Le data linéage aidera à maintenir la qualité des données dans la durée, en organisant la distinction entre processus fournisseurs et consommateurs.

Tagging des données

La documentation des données (structurées ou non) d'une entreprise ajoute de la valeur à travers les tags de ces dernières. Ces systèmes de tagging permettent de gérer les métadonnées et d'accroître la connaissance de l'organisation sur les données qu'elle manipule. Cette gestion des tags regroupe une diversité d'actions : **les process, l'organisation, les rôles/responsabilités et les technologies**.

Cette gestion des tags permet de :

- Optimiser **le pilotage de la qualité** et éliminer les ambiguïtés sur les datas,
- Fournir **des analyses plus efficaces** et sécuriser l'utilisation,
- Faciliter **la mise en conformité** dans le cadre des différentes réglementations,
- Installer la confiance dans le cadre d'**un projet de partage de données** avec des partenaires (Data Sharing).

Nos convictions :

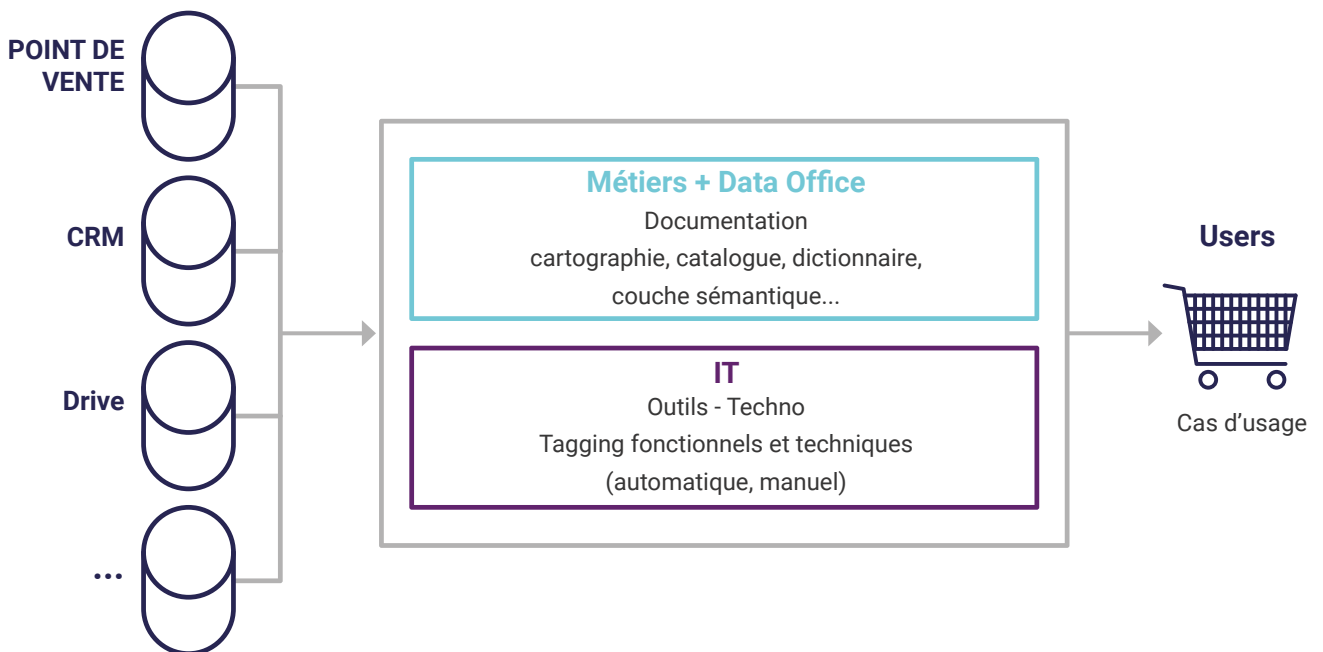
- La mise en œuvre des tags s'insère dans une démarche globale de **documentation des données de l'entreprise** par la création d'une couche sémantique en langage métier qui donne du sens aux données.
- Pour compléter la documentation, les tags sont renforcés par **une cartographie des données** (typologie de données gérée et localisation dans le SI) ainsi qu'un dictionnaire de données faisant le lien entre le niveau technique (Ex. table/traitement) et la couche métier.
- Le facteur clé de succès de la mise en œuvre des tags est **l'implication forte des métiers** pour maintenir une valorisation des données des plateformes data (sensibiliser en continu et responsabiliser).
- Le tagging des données de l'entreprise en un seul lot est à éviter : **il faut construire de manière itérative par périmètre**.
- Pour ce chantier, **la classification des données est un pré-requis important et un processus pivot**. Les Data Officers et métiers fournissent les règles et les catégories nécessaires aux résultats de la classification. La classification est aussi dépendante de l'existence de politiques (RGPD, Risque, Compliance...) qu'il est aussi nécessaire de codifier. Le processus de classification peut être implémenté à plusieurs étapes : collecte des données, création d'une information (agrégats, indicateurs...), contrôle qualité, etc.

Nos convictions :

- Il est important de construire un « **Tagging By Design** » de la data depuis la collecte, le cycle de vie et le management des données :
 - **Le bloc documentation** adressé par les métiers. Il a la charge d'écrire et de maintenir la documentation et de définir les rôles et responsabilité dans la maintenance des données.
 - **Le bloc Tagging**, basé sur des outils, adressé par les métiers et par l'IT. Il permet le Tagging fonctionnel et technique (essentiellement automatique) aboutissant à la couche sémantique partagée par toute l'entreprise et adoptant un langage commun entre les utilisateurs métier des données.
- Les outils ne sont pas une fin en soi dans un tel projet mais ils permettent d'automatiser l'industrialisation de la documentation des données de l'entreprise et sa valorisation.
- La plupart des outils de MDM se limitent à la gestion des référentiels et leurs métadonnées (pour certains).

Tagging by design

Données Opérationnelles





Anonymisation des données

La protection des données personnelles

(toute information se rapportant à une personne identifiée ou identifiable) est devenue une exigence réglementaire forte applicable à toute entreprise depuis mai 2018. La mise en conformité obligatoire avec le Règlement Européen sur la Protection des Données (RGPD) évite les éventuelles sanctions

(jusqu'à 20 millions d'euros ou 4% du chiffre d'affaires mondial) et une dégradation du business (par fuite de données ou mauvaise réputation).

Les entreprises ont été amenées à trouver des solutions pour sécuriser les données sensibles qu'elles détiennent.

Nos convictions

- L'anonymisation des données s'appuie sur plusieurs volets : la politique de sécurité de l'entreprise (données de production), la cartographie des données sensibles, le maintien de l'intégrité et de la cohérence des données intra/inter-applications ainsi que les techniques de cryptage des données jugées sensibles.
- L'anonymisation des données doit être définie par leur usage :
 - ▶ **Protection du stockage des données** : les données sont exportées de la production vers des environnements hors production pour des besoins de test (développement, intégration, recette, formation, reproduction d'incidents, archives, etc).
 - ▶ **Protection de la diffusion des rapports analytiques** : les données sont exportées de la production vers un Data Lake ou des Data Warehouses destinés uniquement à des analyses statistiques sur des populations ou sur des comportements, et non sur des individus spécifiques.
 - ▶ **Sécurisation de la transmission des données** :
Les données sont exportées de la production vers des tiers partenaires ne nécessitant pas l'identification des individus, ou rendues publiques en Open Data ou dans le cadre d'un concours ou d'un programme de recherche avec un contrôle d'accès aux données (qui, quoi, pourquoi).
- L'anonymisation conserve les caractéristiques des données : cohérence des champs de jointure, valeurs anonymisées fonctionnelles (pas de XXX), conservation du format de la donnée initiale...
- **Réversibilité de l'anonymisation** : l'efficacité de l'anonymisation des données dépend de l'identification correspondant soit à l'action d'établissement de l'identité d'un individu, soit de la reconnaissance d'un élément de signalement ou d'information comme appartenant à un individu. Si les données ne sont plus identifiables, les règles en matière de protection des données ne s'appliquent plus.

Nos convictions

- Le choix des méthodes d'anonymisation doit être fait selon le cas d'usage : appauvrissement, hashage, chiffrement, masquage, obstruction, remplacement aléatoire, remplacement pré-calculé, suppression, variance, vieillissement, etc.
- La prise en compte des **impacts organisationnels** est souvent négligée : signature d'accords de non-divulgence, accès sécurisé aux données confidentielles et sensibles, application des règles minimales d'anonymisation, adéquation des processus métiers et procédures organisationnelles.
- Une solution d'anonymisation efficace est une **solution exécutable en « temps réel »** et répondant à 3 contraintes majeures :
 - Elle devra être la plus rapide pour ne pas freiner la propagation des données dans les différents systèmes en quasi-temps réel,
 - Elle devra être la moins destructrice possible d'information,
 - Elle devra optimiser l'équation coût performance.

Le choix des méthodes d'anonymisation doit être fait selon le cas d'usage.



Fournisseurs de données

Les entreprises en quête de performance, d'innovation et de croissance se focalisent toutes sur l'exploitation des données mais pour tirer parti des socles data, elles doivent impérativement utiliser, en complément de leurs données internes, des données externes.

La combinaison et le croisement de ces sources permettront **le développement du business** et le succès de la conquête commerciale.

Pour réussir cette opération, une appréciation de la valeur du fournisseur de données est importante pour mesurer sa couverture et sa complétude, la profondeur et la véracité des données ainsi que sa pérennité.

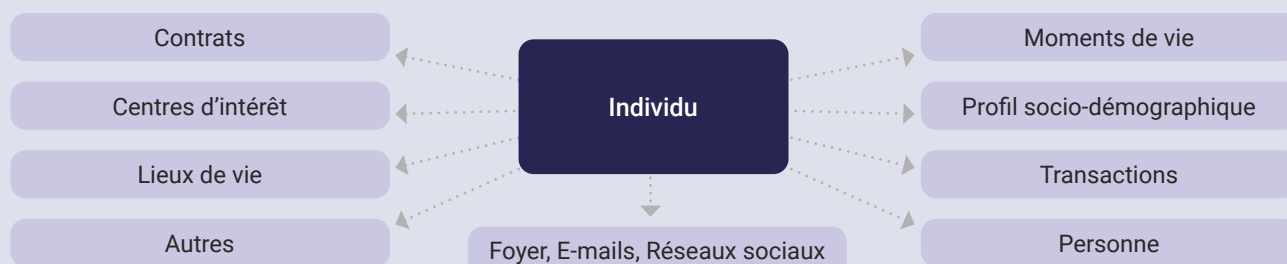
Une cartographie de la valeur data externe est également à réaliser pour mettre en place une stratégie de sourcing transverse à toute l'entreprise.

Stratégie de partage de données avec des fournisseurs tiers

Demande client

Achat de leads / Location de bases de données Enrichissement des données existantes

Data disponibles chez le fournisseur



Partenaires Data du fournisseur

Bases contacts

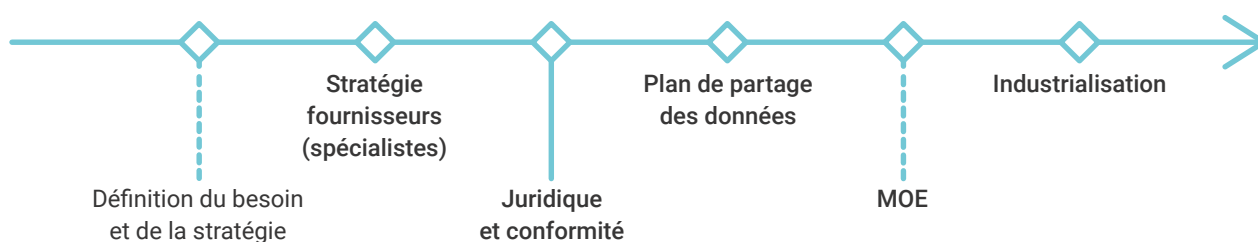
Bases comportementales

Bases contextuelles

Nos convictions

- Les fournisseurs proposant de la donnée **gratuitement** sont souvent inaptes à maintenir un niveau de qualité dans la durée et de manière industrielle.
- En stratégie d'achat, **il faut différencier le fournisseur** par :
 - Son aptitude à identifier les tiers auxquels se rattachent les données et s'assurer que toute nouvelle mise à jour de donnée se raccroche au référentiel maître (ossature),
 - Sa capacité à s'interfacer nativement avec un système de MDM du marché,
 - Le degré de sectorisation couvert par ce dernier,
 - Les références projets correspondants à vos besoins métiers,
 - Les techniques et les approches proposées pour la collecte et la restitution de la donnée : analytiques en temps réel, géo-marketing, études, recherches, analytics vidéos et audios...
 - Le niveau d'insight fourni : convictions, idées...
 - Le modèle de facturation et du business modèle du fournisseur : freemium, à l'usage, par abonnement, combiné...
 - Le modèle de livraison des informations : simple, smart, incluant des données re-traitées, adaptatif lié à la requête, données brutes, via un portail, via des API, via une marketplace...
- Il faut assurer **l'implication des équipes en charge de la mise en conformité, du juridique et des risques** pour mesurer l'impact de l'utilisation des données et leur propagation dans les systèmes d'information (localement, pour des filiales ou à l'international).
- **Définir des SLAs détaillés** intégrant des clauses liées à :
 - La consolidation des fournisseurs de données dont le marché devrait se concentrer. En effet, peu d'entre eux peuvent se maintenir face à l'explosion des coûts liés à la demande d'information croissante et transverse, les technologies de diffusion et d'analyse, la qualité des données ou la fourniture de métadonnées.
 - La provenance des données et leur historique de construction.
 - La certification de ces données.
 - Le droit d'usage des données et des informations y compris celles construites à partir des données utilisées, ainsi que la responsabilité liée à cet usage.
- Pour maintenir un niveau d'apport de valeur dans la durée, **une gouvernance des données externes** est à mettre en œuvre sur les aspects organisation, sécurité, cycle de vie et qualité.

Méthodologie d'accompagnement du sourcing de données externes



Organisation des données

Pendant 20 ans, les **Data Warehouses** ont imposé leur organisation des données :

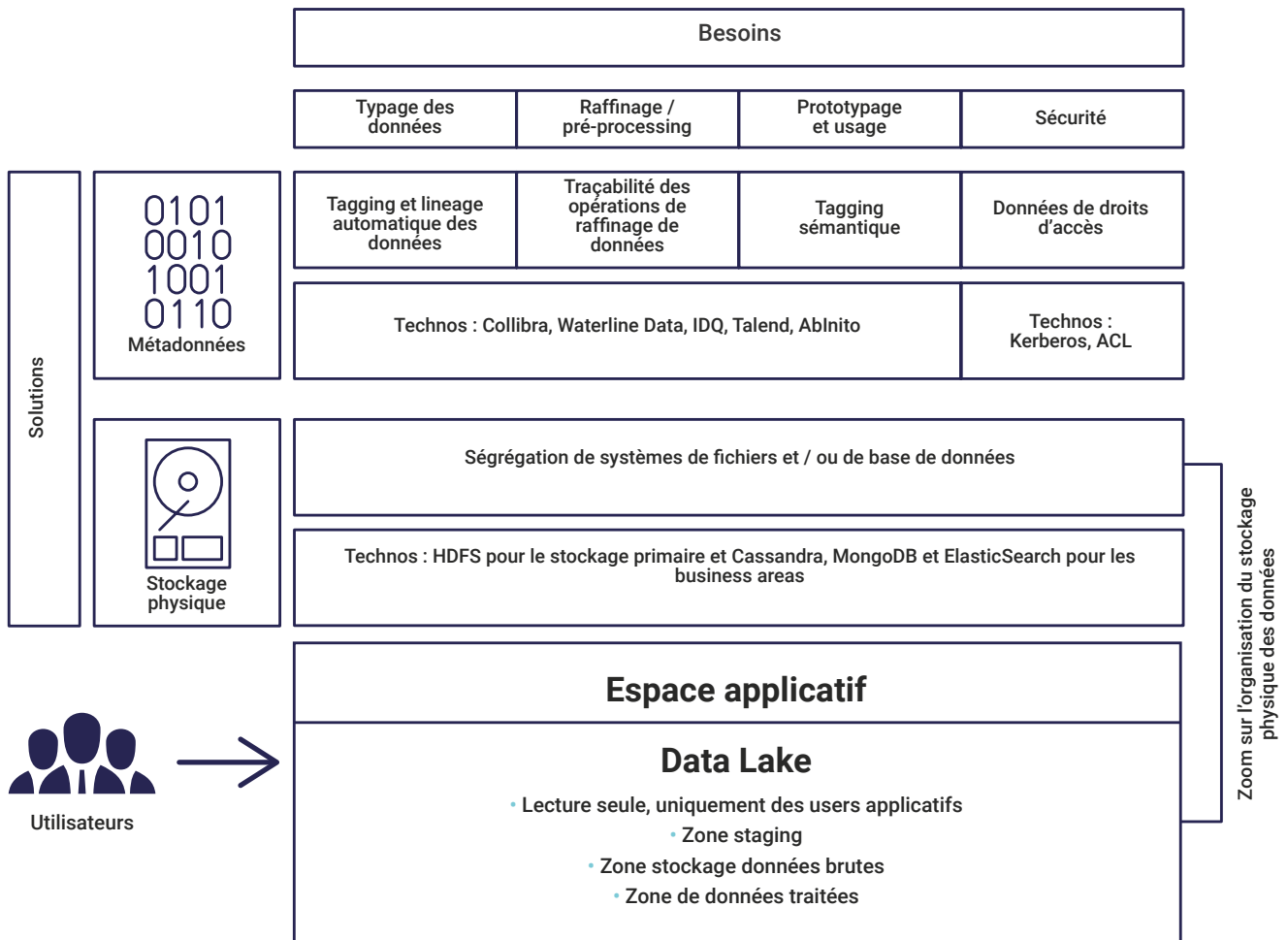
- Un Operational Data Store ou ODS (copie intégrale de données de la production ou delta de production),
- Des espaces de données de références techniques,
- Un espace de données élémentaires et agrégats,
- Des espaces de données historiques.

A ces décompositions du Data Warehouse, s'ajoutaient des **Datamart** (regroupements fonctionnels de données destinés à un métier/fonction de l'entreprise comme par exemple le marketing). Ce passage d'une organisation de données par silo à un socle data & IA partagé et alimenté en temps réel par tous types de données (externe, interne structurée ou non) impose une autre organisation, pierre angulaire de la réussite de toute stratégie data.

Nos convictions :

- Le principe généralement adopté pour gérer et intégrer dans la durée **de forts volumes de données** est de combiner l'utilisation d'une gestion du cycle de vie de la donnée en fonction de sa spécificité.
- Au final et selon notre expérience, 90% des données à destination d'un Data Lake sont déjà structurées. La vraie nouveauté vient plutôt **du mode « temps réel »**.
- La réussite d'un projet de plateforme tient à **la définition préalable d'un modèle d'organisation des données**, matérialisé par des métadonnées et des modalités physiques, en amont des travaux de développement.
- Une approche pragmatique de mise en œuvre repose sur l'intégration de nouvelles données dans ce dispositif **au fil de la réalisation de nouveaux cas d'usage**.
- **Une maintenance continue de cette organisation** est nécessaire pour éviter la péremption de données sur la plateforme et l'apparition d'un nouveau SI Legacy, avec des coûts importants de maintenance ou de stockage de données dupliquées.
- **Un environnement de prototypage ségrégué**, intégré dans la plateforme disposant de l'accès à une large gamme de données et d'un dispositif de sécurité spécifique est un accélérateur pour la mise en œuvre et l'industrialisation des cas d'usage.

Quelle organisation des données mettre en place ?



Glossaire

- **Agile (Méthodes)** : Les méthodes agiles sont des pratiques de pilotage et de réalisation de projets qui reposent sur une approche itérative, incrémentale et adaptative.
- **Algorithme** : Ensemble de règles opératoires dont l'application permet de résoudre un problème énoncé au moyen d'un nombre « n » d'opérations. Un algorithme peut être traduit, grâce à un langage de programmation, en un programme exécutable par un ordinateur.
- **Analyse perspective** : L'analyse perspective permet de synthétiser un grand nombre de données internes et externes pour réaliser des prédictions et suggérer des scénarii de décisions afin de répondre au mieux aux prédictions.
- **Analyse prédictive** : L'analyse prédictive est définie comme l'analyse de données historiques et actuelles disponibles sur un client afin de créer des prévisions sur ses comportements, préférences et besoins futurs.
- **API** : Acronyme pour Applications Programming Interface. Une API est une interface de programmation qui permet de se connecter sur une application afin d'échanger des données. Une API est dite ouverte lorsque le propriétaire du programme le met à disposition publiquement. Les API sont utilisées dans différents domaines du marketing digital.
- **Big Data** : Concept de gestion des données intégrant la capture, la recherche, le partage, le stockage, l'analyse et la présentation des données.
- **Business Model** : Modèle de fonctionnement de l'entreprise, façon dont elle opère et crée de la valeur afin d'assurer sa pérennité.
- **Chaîne de valeur** : Etude précise des activités de l'entreprise afin de mettre en évidence ses activités clés, c'est-à-dire celles qui ont un impact réel en matière de coût ou de qualité et qui lui donneront un avantage concurrentiel.
- **Conduite du changement** : La conduite du changement consiste à anticiper les risques relatifs à un changement, puis à définir et mettre en œuvre une démarche permettant sa mise en place dans des conditions optimales.
- **CRM (Customer Relationship Management)** : Traduit en français par Gestion de la Relation Client, il s'agit de l'ensemble des dispositifs, opérations marketing ou de support ayant pour but d'optimiser la qualité de la relation client, de fidéliser la clientèle et de maximiser le chiffre d'affaires ou la marge par client.
- **Customer Centric** : Consiste à prendre en compte l'impact de chaque décision sur ses clients.

- **Data Crunching** : Le Data Crunching consiste à extraire, à partir de données statistiques désordonnées, des éléments ordonnés propres à former un contenu pertinent en dessinant une cible de prospection et ses besoins d'un côté, et une offre adaptée à ces « requêtes » de l'autre.
- **Data-Centric Company** : Organisation dont le modèle économique repose sur la connaissance et l'analyse de données de ses clients et de son écosystème.
- **Data Lake** : Un Data Lake est un référentiel de stockage qui conserve une grande quantité de données brutes dans leur format natif jusqu'à qu'elles soient exploitées.
- **Data Management Platform (DMP)** : Les plateformes de gestion de l'audience (traduction française) permettent d'accéder aux données de profil d'internautes anonymes afin d'améliorer le ciblage et segmenter cette audience dans un but d'optimisation des actions marketing.
- **DevOps** : DevOps est un mouvement visant à aligner l'ensemble des équipes informatiques, composé de la réunion des « dev » (development engineers) chargés de faire évoluer le système d'information et des « ops » (opérations engineers) chargés d'exploiter les applications existantes au sein d'une même équipe. DevOps vise en particulier à faciliter les évolutions et mises en production de nouvelles fonctionnalités et applications tout en fiabilisant leur fonctionnement.
- **ERP** : Entreprise Ressource Planning (ou Progiciel de Gestion Intégré) est un système d'information destiné à gérer l'ensemble des processus opérationnels d'une entreprise.
- **Expérience client** : Désigne l'ensemble des émotions et sentiments ressentis par un client avant, pendant et après l'achat d'un produit ou service. C'est le résultat de l'ensemble des interactions qu'un client peut avoir avec la marque ou l'entreprise.
- **Internet of Things (IoT) - Internet des Objets (Objets connectés)** : Représente l'extension d'Internet à des objets connectés et à des lieux du monde physique, considéré comme la 3^e révolution d'internet, baptisé Web 3.0.
- **IT** : Information Technology.
- **Itératif (Développement ou méthode)** : Le développement itératif implique de découper un projet en un certain nombre de cycles, ou itérations, au cours desquelles on prévoit de répéter les mêmes activités.
- **Legacy** : Un système hérité, système patrimonial ou legacy system en anglais est un matériel et/ou logiciel continuant d'être utilisé dans une organisation (entreprise ou administration) alors qu'il est supplanté par des systèmes plus modernes. L'obsolescence de ces systèmes et leur criticité les rendent difficilement remplaçables sans engendrer des projets coûteux et risqués.
- **MDM** : Master Data Management ou référentiel de données.
- **Méthode Kanban** : Définit une méthode simple, visuelle et facilement compréhensible par tous, basée sur le principe du « juste-à-temps ». Ce sont les commandes émises par un poste aval en fonction de ses besoins qui régulent la production d'un poste en amont.

- **Méthode Scrum** : La plus connue des méthodes « agiles » permet une gestion réactive, incrémentale et itérative des différents projets de l'entreprise. Les projets sont divisés en plusieurs cycles de travail.
- **Open Data** : Démarche de publication de données numériques en ligne selon une série de 10 critères garantissant leur libre accès et leur réutilisation par tous.
- **Open innovation** : Processus par lequel une entreprise est capable de faire appel à des idées et expertises en dehors de ses propres murs. Il permet également à une entreprise ou organisation de rentabiliser ses idées ou brevets en dehors de son propre marché, en les proposant à d'autres entreprises ou institutions.
- **PoC (Proof of Concept)** : Désigne le fait d'avoir des éléments plus ou moins tangibles assurant qu'un nouveau concept de produit ou service n'est pas une simple vue de l'esprit. Le PoC vise à limiter le risque de pertes financières liées à des développements de nouveaux produits ou services (coûts directs de développement et production, temps perdu, etc).
- **Pure Player** : Désigne un acteur exerçant son activité commerciale uniquement sur internet. On oppose généralement les « pure players » et les « click and mortar » qui déploient des activités hybrides.
- **ROI** : Return On Investment (Retour sur Investissement).
- **Scalabilité** : La scalabilité est la capacité d'une entreprise à adapter son business model, son schéma d'entreprise à une forte augmentation de son volume d'activité. Concrètement, sa capacité à passer de la petite startup à la multinationale. La scalabilité permet d'absorber le choc de croissance extrême que certaines startups connaissent.
- **SI** : Système d'Information.
- **Temps réel** : On parle d'un système temps réel lorsque ce système est capable de contrôler (ou piloter) un procédé physique à une vitesse adaptée à l'évolution du procédé contrôlé.
- **Time-to-Market** : Le délai de mise sur le marché, ou Time-to-Market, est le temps moyen écoulé entre la génération d'une idée et sa commercialisation. Il permet de mesurer la capacité de l'entreprise à mettre en œuvre des innovations.
- **Tracking** : Le tracking internet consiste à observer les réactions et actions d'un utilisateur Internet et de comprendre l'effet de la communication publicitaire sur son attitude et son comportement (on peut par exemple observer le comportement d'un visiteur sur un site Web ou au sein d'une application mobile).
- **UX Design** : Approche de design et de conception d'une interface utilisateur (la meilleure possible).
- **Vision client à 360°** : Pour maîtriser parfaitement sa relation client, il est nécessaire de faire l'inventaire de tous les points de contacts entre l'entreprise et sa clientèle, de façon exhaustive et dans la durée. La vision 360° est donc l'aboutissement de ce postulat, c'est-à-dire la connaissance du client sur tous les canaux de communication.



Et vous ?

Vous souhaitez repenser votre Data Management ou établir une stratégie data complète ? Nos experts sont à votre écoute.

DataValue Consulting est un cabinet de conseil et d'intégration indépendant spécialisé dans l'exploitation et la valorisation de la data. Nous accompagnons plus d'une centaine d'organisations publiques et privées dans la définition, la mise en œuvre et l'exploitation de leur patrimoine de données.

Nos 250 consultants passionnés mettent leur expertise technologique et leur expérience en management à disposition de vos ambitions. Nous délivrons un accompagnement à 360° de la définition de roadmap et de cas d'usage, au pilotage de projets de data management, Business Intelligence, data gouvernance, data visualisation, architecture, data science et pilotage de la performance (EPM).

Rédacteurs et contributeurs



● **Abdelaziz Joudar**
Président du groupe



● **Thierry Deliens**
Directeur Avant-Vente



● **Refaï El Khattabi**
Directeur BU Data

Contact

● **Contact Commercial** : contact@datavalue-consulting.com



DataValue Consulting

9 Boulevard du Général de Gaulle, 92120 Montrouge
Tél : 01 81 89 41 70 - www.datavalue-consulting.com
contact@datavalue-consulting.com

