



Gouvernance et Architecture Data & Analytics

Élaborer et mettre en œuvre la stratégie data

Janvier 2023



Cigref

Gouvernance et Architecture Data & Analytics

Élaborer et mettre en œuvre la stratégie data

Janvier 2023



Droit de propriété intellectuelle

Toutes les publications du Cigref sont mises gratuitement à la disposition du plus grand nombre mais restent protégées par les lois en vigueur sur la propriété intellectuelle

SYNTHÈSE

Les entreprises et administrations publiques sont convaincues de la valeur de la donnée. Les données constituent en effet, un actif propre pour l'entreprise entière et pas seulement pour un seul service ou une direction métier. C'est pourquoi les entreprises et les administrations publiques se transforment afin de mieux valoriser ces données et investissent dans la construction et la mise en place d'une stratégie data au service de leur stratégie globale.

Pour élaborer leur stratégie data, les organisations identifient leurs motivations, les attendus de cette démarche et évaluent les gains escomptés : meilleur *time-to-market*, conquête de nouveaux marchés, excellence opérationnelle, amélioration et ajustement des réponses aux clients/utilisateurs et garantie de la conformité des solutions et produits. Le contexte propre de chaque organisation permet de placer la stratégie data dans le cadre d'une stratégie globale qui répond à sa raison d'être ainsi qu'à des besoins structurels et/ou conjoncturels (organisation du groupe, crise sanitaire, etc.).

Ensuite, la construction de la stratégie data se décline en quelques grandes étapes, qui peuvent être appréhendées avec la démarche suivante :

1. Évaluer la maturité de l'entreprise et développer la culture des données des collaborateurs ;
2. Définir des objectifs métiers précis ;
3. Identifier les données pertinentes à utiliser ;
4. Préparer l'architecture et les briques techniques.

Ce « mode d'emploi » de la construction de la stratégie data est la synthèse des diverses approches des contributeurs au groupe de travail « Gouvernance et architecture data ». Il est bien entendu susceptible d'être amendé et adapté à chaque organisation pour répondre à ses besoins propres.

Une fois la stratégie data définie, il faut l'implémenter au niveau des fondations techniques, de la gouvernance, de l'organisation et des processus. L'architecture du Système d'Information (SI), son historique, les dépendances technologiques et la maturité de la culture data sont autant d'éléments qui modulent cette implémentation.

La stratégie data impose à l'entreprise de se doter d'une gouvernance data dédiée, afin de répondre à des enjeux tels que la connaissance de la data, la maîtrise de la qualité, l'harmonisation des données de référence, la conformité réglementaire, etc. Tous les acteurs devant œuvrer ensemble pour préparer les données à des usages, il est important de bien articuler la gouvernance de la stratégie data avec les autres gouvernances mises en place, comme celles de la stratégie des SI, de la stratégie des métiers et de la transformation digitale qui concourent toutes à la stratégie générale de l'entreprise.

Pour parvenir à la mise en œuvre de la stratégie data, les entreprises et les administrations publiques travaillent en outre sur son organisation, à savoir la répartition des missions entre les collaborateurs et la définition du rôle de chacun dans le déploiement de cette stratégie. Cette question est d'ailleurs de plus en plus abordée dans le cadre de l'entreprise agile.

Enfin, une fois la stratégie data définie et mise en place, reste pour l'entreprise ou l'administration publique le soin de piloter son implémentation, de suivre sa mise en œuvre et d'évaluer ses bénéfices sur différents axes. Le rapport partage les indicateurs et mesures utilisés à ces fins par les participants du groupe de travail « Gouvernance et architecture data ».

REMERCIEMENTS

Nos remerciements vont à **Alice Guéhennec**, *Chief Digital & Information Officer* du groupe SAUR et à **Patrick Mahu**, architecte d'Entreprise chez Pôle Emploi qui ont piloté ce travail, ainsi qu'à toutes les personnes qui ont participé et contribué à ce groupe de travail Cigref (ordre alphabétique) :

Mourad ABBAS – RENAULT	Stéphane – GUILLOTVALLOUREC
Olivier ACCOLAS – GROUPE EGIS	Jocelyn HABEILLON-WOLF – HAGER GROUP
Emmanuel BATT – GROUPE SEB	François HERLENT – MAIF
Sindbad BELLILI – SYSTÈME U	Alexandre HUDELLOT – SCOR
Frédéric BERNARD-PAYEN – AIRBUS	Cédric JUBLOT – EIFFAGE
Frédéric BERQUEZ – MATMUT	Christophe JUILLET – EDF
David BIREBENT – LABORATOIRES PIERRE FABRE	David KRIEFF – GROUPE ADP
Mohamed BOUTAYBI – BA1NQUE DE FRANCE	Andreea LACHAPELLE – EDENRED
David BRETON – FAYAT	Raynald LASOTA – FRANCE TELEVISIONS
Léa CAPGEN – SCOR	Rudy LEROUX – BANQUE DE FRANCE
Laurent CASSASSOLLES – RENAULT	Patrick MAHU – PÔLE EMPLOI
Xavier CASTA – POLE EMPLOI	Roland MANFREDI – GROUPE ADP
Evangelos CHAIDACIS – TOTALENERGIES	Philippe MORILLON – KEOLIS
Florence CHIRAT – GROUPE SAVENCIA	Lionel NAESSENS – MSA
Raphaël COMMARET – VIRBAC	Sandra NAUD – SYSTÈME U
Nicolas CONNAULT – NAVAL GROUP	Valérie PENAUD – POLE EMPLOI
Frédéric DAMEZ – ESSILOR	Pierre RAMEL – BNP PARIBAS
Aurélie DAVIAUD – GROUPE BPCE	Franck RICOTTIER – AIR FRANCE KLM
Xavier DELAGE – POLE EMPLOI	Hugues ROUMEZIN – MAIF
Stéphanie DELAUNAY – ENEDIS	David SCEMAMA – MINISTERE DE L'INTERIEUR
Fabrice DESCHAMPS – COVEA	Ronan TIGREAT – NAVAL GROUP
Gregory DENIS – SAUR	Gilles VAN SCHENDEL – GROUPE SAVENCIA
Hassan DRISS – MINISTERE DE L'INTERIEUR	Claude VARY – GROUPE ADP
Bertrand DUBRULLE – CNAV	Florent VERRIERE – EDF
Youssef EL MOUDDEN – SCOR	Dominique VIENNE – ENEDIS
Stéphane GIRARD – TDF	Frédéric WALKER – MSA
Alice GUEHENNEC – SAUR	

Nous remercions également vivement tous les intervenants qui ont apporté beaucoup de matière à la réflexion de notre groupe de travail (ordre alphabétique) :

- Sindbad BELLILI, Responsable de l'équipe EDH, Système U ;
- Frédéric BERNARD-PAYEN, Data Governance Manager, Airbus ;
- Gregory DENIS, Chief Data Officer, SAUR ;
- Sylvain GIRARD, Chief Digital & Information Officer, Hager Group ;
- Alice GUÉHENNEC, Chief Digital & Information Officer, SAUR ;
- Jocelyn HABEILLON-WOLF, Data Management Specialist, Hager Group ;
- Christophe JUILLET, Responsable de l'architecture d'entreprise, EDF ;
- Frédéric LAU, Directeur de mission, Cigref.
- Marie-Emmanuelle LECLERCQ, Data Governance Manager, SAUR ;
- Patrick MAHU, Architecte d'entreprise, Pôle Emploi ;
- Lionel NAESSANS, Data Architect, iMSA ;
- Hugues ROUMEZIN, responsable de la Gouvernance de la Data et de la qualité des données, MAIF ;

Ce document a été construit et rédigé par **Marine de SURY**, directrice de mission au Cigref avec le support d'**Elena SILVERA**, chargée de mission au Cigref.

TABLE DES MATIÈRES

1 INTRODUCTION	6
2 CONSTRUCTION DE LA STRATÉGIE DATA	7
2.1 Pourquoi élaborer une stratégie ?	7
2.1.1 Conquête de marchés.....	8
2.1.2 Défense contre le risque extérieur d’innovation de rupture.....	8
2.1.3 Meilleur ajustement des réponses aux clients/utilisateurs des services publics	9
2.1.4 Amélioration de l’excellence opérationnelle.....	9
2.1.5 Conformité, réglementation, qualité.....	10
2.2 Élaborer une stratégie data	10
2.2.1 Étape 1 : Évaluer la maturité de l’entreprise et développer la culture data des collaborateurs	11
2.2.2 Étape 2 : Définir des objectifs métiers précis	15
2.2.3 Étape 3 : Identifier les données pertinentes à utiliser.....	17
2.2.4 Étape 4 : Préparer l’architecture et les briques techniques	18
3 MISE EN OEUVRE DE LA STRATÉGIE DATA	19
3.1 Fondations techniques.....	19
3.1.1 Captation & Stockage des données	20
3.1.2 Briques de Data gouvernance.....	21
3.1.3 Exploration, exploitation et Business intelligence.....	23
3.1.4 Stratégie logicielle et infrastructure	23
3.2 Gouvernance	24
3.2.1 Connaître les données (processus « KYD Know Your Data »).....	25
3.2.2 Maîtriser la qualité des données	26
3.2.3 Conformité réglementaire & protection.....	26
3.3 Organisation.....	26
3.3.1 Fédération plutôt que décentralisation.....	27
3.3.2 Contrôle et responsabilité de la donnée	27
3.3.3 Processus	28
3.4 Piloter l’implémentation de la stratégie data et suivre ses bénéfices.....	28
3.4.1 Piloter la mise en œuvre de la stratégie data.....	29
3.4.2 Mesurer les bénéfices business et la valeur apportés par l’usage de la data	31
4 CONCLUSION	33

TABLE DES FIGURES

Figure 1 Étapes de l'élaboration de la stratégie data - Source Cigref.....	11
Figure 2 Briques communes dans les fondations de l'architecture pour déployer la stratégie data	19

TABLE DES ENCARTS

Retour d'expérience EDF : Acculturation et formation.....	12
Retour d'expérience Cnav : Acculturation à la qualité des données.....	14
Retour d'expérience Hager Group : Exemple de leviers de la construction de leur stratégie data ..	15
Retour d'expérience du groupe SAUR : Exemple de déclencheur de la stratégie data.....	16
Retour d'expérience de Pôle Emploi : Exemple de déclencheur de la stratégie data	16

1 INTRODUCTION

Face au raz de marée des données issues d'origines diverses, capteurs, machines, systèmes complexes, experts humains, environnement, etc., se pose la question de leur valorisation à travers des méthodes, des démarches et des processus, dans des contextes très divers : données de gestion, données des clients, données scientifiques ou industrielles, open data¹, auxquelles peuvent être ajoutées des métadonnées permettant de les caractériser.

En 2016, le groupe de travail Cigref « [valorisation des données dans les grandes entreprises](#) »² s'était penché sur les meilleures pratiques de valorisation des données et avait élaboré un [outil d'auto-évaluation](#)³ de la maturité d'une organisation en matière de valorisation des données. Il ressort de ces travaux que la donnée se doit d'être comprise et gérée comme un **actif propre de l'entreprise ou de l'administration publique**, dont la propriété n'est pas exclusive à une population ou une entité particulière, **tous les acteurs devant œuvrer ensemble pour préparer les données à des usages**.

Alors qu'il est souvent affirmé que « les données sont capitales dans l'entreprise », le passage d'une aspiration du top management à une mise en œuvre de la stratégie data est moins évident. Ce rapport a pour objectif de fournir à la direction des systèmes d'information et à la direction du digital, les clés pour parvenir à répondre aux attentes fortes de la direction générale. Il s'adresse à l'ensemble des collaborateurs qui sont appelés à devenir parties prenantes de la stratégie data.

Les entreprises et administrations publiques se transforment afin de mieux valoriser les données et cela constitue une opportunité pour servir la stratégie de l'entreprise à travers la stratégie data. La première partie du rapport vise à répondre à deux interrogations : « pourquoi mettre en place une stratégie data ? » d'une part, et d'autre part « pour quoi ? », « dans quel(s) but(s) ? ». Il convient de préciser à ce stade que la valorisation de la donnée ne peut pas être conçue comme un sujet purement technologique. Pour cause, **la stratégie data est intrinsèquement liée à la raison d'être** de l'entreprise ou de l'administration publique, à laquelle elle doit répondre. Cette partie propose également quelques grandes étapes pour élaborer et construire la stratégie data de l'entreprise ou de l'administration publique.

La stratégie data touche à l'image-même de l'organisation, de ses valeurs, de ses aspirations et projets pour l'avenir : c'est un sujet de transformation d'entreprise. Le rapport propose de bonnes pratiques pour sa mise en œuvre. La bonne démarche est celle qui permet d'atteindre les objectifs définis dans la stratégie, sur des sujets comme la structuration du programme de transformation dans l'entreprise, les aspects de plateforme technologique, la gouvernance, la gestion du portefeuille des projets/produits et services, ou bien l'industrialisation.

Il est apparu que l'identification claire des objectifs de la stratégie data devait être tout particulièrement soignée par l'entreprise de façon à mesurer les résultats afin de les comparer aux objectifs attendus. Aussi, la mesure pertinente des résultats facilite l'organisation de cette stratégie et permet d'en planifier les étapes, de déployer utilement les ressources, et de « mener les chantiers dans le bon ordre ». Bien conscient des enjeux qui sous-tendent la gouvernance et l'architecture data, le groupe de travail « Gouvernance et architecture de la data » s'est donné pour objectifs de répondre à ces deux problématiques principales, et de mesurer les bénéfices du bon pilotage de la politique des données.

¹ En français, données ouvertes : les données ouvertes-sont disponibles en libre accès et peuvent être utilisées et partagées librement.

² <https://www.cigref.fr/un-rapport-cigref-sur-la-valorisation-des-donnees-en-entreprise>

³ <https://www.cigref.fr/un-rapport-cigref-sur-la-valorisation-des-donnees-en-entreprise>

2 CONSTRUCTION DE LA STRATÉGIE DATA

Les entreprises et administrations publiques sont convaincues de la valeur portée par les données. Ces dernières représentent un **actif de l'entreprise**, et n'appartiennent pas seulement à un service ou à une direction métier. Il est donc important que la gouvernance de la donnée soit organisée **au niveau de l'entreprise** avec un **responsable**, qui veille à sa qualité, la gère en fonction de son cycle de vie, la sécurise, la rende disponible selon des règles à préciser et connaisse ses différentes utilisations en interne ou auprès de son écosystème (fournisseurs, partenaires, clients, vendeurs, etc.).

Tout le monde s'accorde à dire que la donnée est l'or noir d'aujourd'hui, bien que celle-ci soit un "bien non rival". En effet elle possède une valeur qui peut se décliner de trois façons : valeur d'usage liée à son utilisation (plus la donnée publiée est utilisée plus sa valeur croît) ; une valeur de marché liée à sa rareté ; et enfin une valeur de protection liée à sa sensibilité.

La maturité quant à la valorisation des données est très variable selon les entreprises et administrations publiques. Certaines ont déjà défini la stratégie data et sont en train de la déployer. La data sert leurs objectifs stratégiques. En effet, **la stratégie data est avant tout une stratégie d'entreprise**. Encore **faut-il que la donnée soit bien documentée, identifiée, à jour et exploitable**. D'autres en sont au stade de la prise de conscience et commencent à réfléchir à la façon de valoriser la donnée. D'autres encore, cherchent à adopter une vision plus « data centrée » en leur sein, c'est-à-dire d'intégrer la donnée dans la prise de décision. La question de l'intérêt d'une stratégie data se pose donc pour plusieurs d'entre-elles.

Après avoir répondu au *pourquoi* élaborer une stratégie data, ce chapitre étudie *comment* la construire et propose quelques bonnes pratiques pour son élaboration.

2.1 POURQUOI ÉLABORER UNE STRATÉGIE ?

L'enjeu dans la construction d'une stratégie data pour l'entreprise ou l'administration publique est bien de **créer de la valeur** pour les directions métiers et également plus globalement, pour l'ensemble des équipes. Plusieurs objectifs sont poursuivis à travers la valorisation de la donnée, et l'élaboration d'une stratégie data est généralement induite par l'un de ceux-ci, listés ci-après :

- La conquête de marchés
 - Amélioration du service client ;
 - Vente de services ;
 - Gestion du risque.
- La défense contre le risque d'innovation de rupture, externe à l'entreprise ou l'administration publique, utilisant la donnée mise à disposition (données ouvertes ou open data).
- Pour les services publics, l'amélioration en qualité et pertinence des réponses aux clients/utilisateurs
 - Réponses adaptées aux nombreux clients
 - Lutte contre le non recours aux droits et lutte contre la fraude des services publics
- L'amélioration de l'excellence opérationnelle

- Amélioration de l'efficacité opérationnelle ;
 - Mise à disposition, usage et maîtrise par le client final ;
 - Meilleure adéquation des réponses aux clients du service public.
- Conformité / Réglementation / Qualité
 - Prise en compte des aspects environnementaux et développement durable ;
 - Responsabilité sociale
 - Conformité au Règlement Général sur la Protection des Données (RGPD).

2.1.1 CONQUÊTE DE MARCHÉS

Souvent, des projets exploitant les données existent déjà au sein de l'entreprise ou de l'administration publique. **Fédérer ces initiatives** permet parfois de développer de nouvelles prestations de service qui offrent des opportunités *business* en sus des activités traditionnelles de l'entreprise. Par exemple via la vente de données collectées ou générées, ou via celle de services créés par leur exploitation.

Plusieurs entreprises ou administrations publiques ont cherché à évaluer la chaîne de valeur de leur parcours client/utilisateur, ce qui leur a permis d'identifier les points d'amélioration de leurs services en interne et/ou auprès de l'écosystème.

D'autres encore, ont initié la construction de leur stratégie data afin d'affiner la connaissance de leurs clients/utilisateurs (détermination et suivi du profil client avec ses comportements de consommation). Ce ciblage plus précis leur permet d'optimiser les ventes et/ou met en lumière des pistes pour de nouvelles opportunités *business*. Encore faut-il que la stratégie data puisse être mise en œuvre facilement, dans un temps adapté pour répondre au *time to market*, et avec un investissement acceptable.

Pour certains, l'élaboration d'une stratégie data commence avec la volonté de digitaliser des services, ou de les *redesigner* afin d'apporter davantage de valeur aux clients et/ou d'accroître leur monétisation.

Enfin, certaines entreprises ont construit leur stratégie data après l'acquisition de nouvelles entités afin d'intégrer les nouveaux métiers et de tirer des bénéfices rapidement.

2.1.2 DÉFENSE CONTRE LE RISQUE EXTÉRIEUR D'INNOVATION DE RUPTURE

La stratégie data peut être mise en place afin de répondre à l'expression d'une menace externe ou transverse à l'entreprise. En effet, des données publiques de l'entreprise (par exemple, la valeur nutritionnelle de produits alimentaires vendus), peuvent être exploitées par un tiers pour créer un nouveau service (par exemple, [Yuka](#), si nous continuons avec le cas de la valeur nutritionnelle). L'utilisation de ces données peut provoquer alors la déstabilisation de tout un écosystème en constituant une innovation de rupture qui fait évoluer les attentes des clients. Le métier de base est perturbé par l'émergence en périphérie de nouveaux acteurs avec un *business model* basé sur l'exploitation des données du métier initial. Pour l'entreprise qui subit cela, la question est de savoir comment réagir. Les données étaient déjà présentes dans l'entreprise, mais pas utilisées aux mêmes fins. Elles étaient utilisées dans la chaîne d'approvisionnement (*supply chain*) afin de connaître les formules des produits vendus, ou encore en raison de contraintes réglementaires (afficher une formule

sur les étiquettes) et sont devenues un objet de R&D. La spécificité de cette menace externe ou transversale est qu'elle fait évoluer la perception des clients, ce qui oblige l'entreprise à s'adapter pour mieux répondre à leurs attentes. Autre exemple : *Trainline*. *Trainline* est un agrégateur indépendant, qui propose les billets de 270 compagnies ferroviaires, dans quarante-cinq pays européens et qui offre également la possibilité de comparer les concurrents entre eux comme SNCF et *Trenitalia*, entre Paris et Lyon. Cette société britannique qui a intégré le marché français en 2016 en absorbant la jeune pousse tricolore *Captain Train*, bouscule l'offre de SNCF Connect.

Plusieurs startups s'appuient ainsi sur l'*open data*, autrement dit des données ouvertes, mises à disposition par des entreprises ou administrations publiques pour proposer de nouveaux services.

Dans les « *business de plateforme* » les *business* sont proposés à partir du point d'entrée que constitue la plateforme (Cf. le rapport Cigref [Nouvelles stratégies de plateformes](#)⁴ publié en 2019). La plateforme répond à une problématique client dans un écosystème. Des entreprises de l'écosystème peuvent parfois se trouver pratiquement en sous-traitance et, bien qu'elles aient leur propre site internet ou leur propre plateforme, leur présence sur la plateforme en ligne du prescripteur ou du comparateur incontournable est impérative au regard de la part de marché de celui-ci. Par exemple, la plupart des hôtels se doivent d'être présents sur le site incontournable *Booking.com*, même s'ils ont leur propre plateforme de réservation en ligne. L'intermédiation, par *Booking.com* dans cet exemple, est à l'origine d'une perte de marge des entreprises d'hébergement au profit de cette plateforme.

2.1.3 MEILLEUR AJUSTEMENT DES RÉPONSES AUX CLIENTS/UTILISATEURS DES SERVICES PUBLICS

Les services publics cherchent à améliorer leurs offres de services et à mieux ajuster les réponses proposées aux clients/utilisateurs. L'exploitation de données soigneusement identifiées leur permet de mieux détecter la fraude et de faciliter la prise en main de leurs services par des utilisateurs plus réticents au numérique.

Les administrations publiques cherchent aussi à lutter contre le non-recours aux droits et aux services sociaux qui constituent un des axes stratégiques des politiques de prévention et de réduction de la pauvreté. Le numérique peut être un vecteur d'accès simplifié aux services publics ou encore un moyen pour aider à remédier à la situation. En effet, cette lutte s'appuie sur diverses données (statistiques publiques, enquêtes et expériences spécifiques) qui permettent d'identifier, décrire et évaluer le phénomène du non-recours.

2.1.4 AMÉLIORATION DE L'EXCELLENCE OPÉRATIONNELLE

Favoriser l'anticipation, avoir des éléments tangibles pour orienter les équipes dans leur prise de décision ou encore améliorer l'efficacité des interactions au sein de l'écosystème de l'entreprise avec ses clients, ses collaborateurs, ses fournisseurs et partenaires, grâce à la donnée constituent des leviers pour élaborer la stratégie data.

En effet, les entreprises cherchent à améliorer en permanence leurs solutions et services et les processus associés, en renforçant le contrôle des métiers, en réduisant les coûts, en améliorant les

⁴ <https://www.cigref.fr/wp/wp-content/uploads/2019/12/cigref-nouvelles-strategies-plateformes-business-conception-mise-en-oeuvre-decembre-2019.pdf>

processus internes et l'efficacité opérationnelle et en apportant plus de valeur à leurs clients, et parfois également à leurs clients internes (en particulier dans les organisations disposant de Centre de Services Partagés CSP). Les métriques et les indicateurs, et tout ce qui constitue la « data », offrent la possibilité d'affiner les prévisions, d'accélérer les décisions et de mieux cibler les attentes des clients et donc de renforcer la capacité de prise de décision ou d'orientation stratégique selon l'adage « savoir, c'est pouvoir ». Les équipes s'appuient sur les données de mesure ciblées pour valider ou confirmer des perceptions ou retours de clients, et évitent de développer des solutions/modules avec des perceptions erronées. L'image de marque représente pour l'entreprise un enjeu capital. Les données qui l'évaluent lui permettent d'analyser la perception de ses clients et de ses détracteurs afin de déterminer les actions à mener pour atteindre les objectifs fixés. Les données peuvent ainsi contribuer à façonner l'image de marque.

Enfin, la recherche de l'efficacité des interactions au sein de l'écosystème de l'entreprise, avec ses clients, ses collaborateurs, ses fournisseurs et partenaires constitue pour certaines entreprises et administrations publiques un des éléments déclencheurs de l'élaboration d'une stratégie data. Cette construction nécessite une meilleure autonomie des utilisateurs métiers dans l'utilisation de la donnée pour accroître les interactions avec leur écosystème, et assurer des relations de meilleure qualité. En interne, l'information doit être efficacement partagée entre les différents métiers de l'entreprise et disponible à travers un catalogue de données accessible à tous (Cf. paragraphe 3.1.2.)

2.1.5 CONFORMITÉ, RÉGLEMENTATION, QUALITÉ

De nouvelles réglementations, obligations de marché ou de qualité – liées à la conformité au RGPD, au respect du secret médical et à la cybersécurité, par exemple – constituent dans certains domaines d'activité un catalyseur de l'élaboration d'une stratégie data. Ces contraintes sont alors utilisées comme un levier. La traçabilité des données conduit à une meilleure maîtrise tout au long de leur cycle de vie. Pour certaines entreprises et administrations publiques, l'orientation stratégique de la donnée répond à des enjeux climatiques ou environnementaux comme la décarbonation de la stratégie IT ou le verdissement des données.

2.2 ÉLABORER UNE STRATÉGIE DATA

La stratégie data s'inscrit dans la stratégie de l'entreprise ou de l'organisation publique, que ce soit en termes de performance opérationnelle ou de croissance. L'approche pour élaborer la stratégie data diffère en fonction des entités même si des communs émergent. Il ressort des ateliers du groupe de travail « Gouvernance et architecture data » que la stratégie data se construit en différentes étapes présentées ci-après et s'articule autour de plusieurs enjeux *business*, organisationnels, réglementaires et parfois financiers.

- Étape 1 : Évaluer la maturité de l'entreprise et développer la culture des données des collaborateurs ;
- Étape 2 : Définir des objectifs métiers précis ;
- Étape 3 : Identifier les données pertinentes à utiliser ;
- Étape 4 : Préparer l'architecture et les briques techniques.

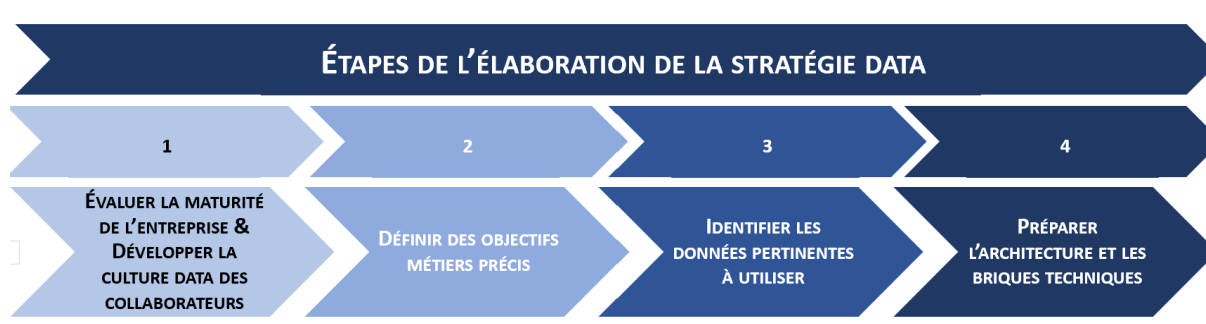


Figure 1 Étapes de l'élaboration de la stratégie data - Source Cigref

2.2.1 ÉTAPE 1 : ÉVALUER LA MATURITÉ DE L'ENTREPRISE ET DÉVELOPPER LA CULTURE DATA DES COLLABORATEURS

2.2.1.1 Évaluer la maturité de l'entreprise concernant les data

Plusieurs contributeurs au groupe de travail « Gouvernance et architecture de la data » préconisent d'établir la stratégie data en prenant en compte la maturité collective de l'entreprise concernant les données. Il s'agit de constater d'où l'entreprise part, en évaluant sa maturité et en partageant cet état des lieux avec les collaborateurs. Cela permet de déterminer les étapes restant à franchir pour atteindre le niveau ciblé. Cette évaluation peut être faite avec l'outil d'auto-évaluation de la valorisation de la donnée du Cigref, publié en 2016 avec le rapport éponyme « [Valorisation des données dans les grandes entreprises](#) ». La grille de maturité est répartie en axes structurants :

1. *Business Models*
2. Connaissance (clients, écosystème)
3. Transversalité et efficacité opérationnelle
4. Stratégie et gouvernance
5. Éthique et conformité
6. Outils et solutions
7. Sobriété numérique

Ce dernier axe structurant de la sobriété numérique a été ajouté à la suite du projet Cigref / Telecom Paris Tech - Master 1, mené par des étudiantes de cursus « Management des Systèmes d'Information et de la Transformation Numérique » effectué dans le cadre du groupe de travail « Gouvernance et architecture data ».

L'évaluation de la maturité peut révéler l'existence, dans une même entreprise, d'écarts d'appréciation entre les différentes équipes. L'outil d'auto-évaluation, de par ses questionnements, assure à l'entreprise ou à l'administration publique de balayer les sujets les plus importants. Il constitue un bon moyen d'échange en interne permettant aux équipes de confronter leurs perceptions respectives par rapport aux enjeux de la stratégie data.

2.2.1.2 Acculturer aux données, former et communiquer

L'objectif de cette sous-partie est de présenter les différentes façons d'effectuer cette transformation des mentalités partagées par les participants du groupe de travail.

Acculturer

Les entreprises n'ont pas forcément conscience de ce qu'elles peuvent tirer du partage et de l'exploitation des données. Cela nécessite d'acculturer et de former les collaborateurs en s'appuyant sur des exemples concrets de création de valeur générée pour les métiers. Il faut aussi aborder le coût induit pour qualifier les données, les stocker, les consolider, voire les détruire. La data doit être appréhendée comme un **outil** (*enabler*).

Pour embarquer les métiers, il peut être pertinent de montrer l'impact de la donnée en s'appuyant sur des cas d'usage concrets, de faire part de retours d'expériences, ou encore d'utiliser des données déjà disponibles (jeux de données ou *datasets*, KPIs⁵ et tableau de bord ou *dashboard*) pour répondre à des enjeux posés par la direction sur des projets existants.

Certains collaborateurs ont parfois une vision des données réduite à celle d'un simple fichier Excel. D'où l'importance d'expliquer la variété et la richesse de ces données qui peuvent être simples ou complexes après traitements d'intelligence artificielle ou analytiques. **Sensibiliser à la qualité de la donnée, à son cycle de vie**, ainsi qu'aux **contraintes légales et aux réglementations locales ou globales** auxquelles les données personnelles (RGPD) et industrielles (*Data act*) sont soumises, constitue un enjeu essentiel auprès des salariés. Afin de pouvoir délivrer un service de confiance aux clients, les données doivent être caractérisées en toute transparence en tenant compte des risques éthiques ou juridiques attachés à leur utilisation. Ceux-ci peuvent porter par exemples sur leur origine, si elles proviennent de l'écosystème (données externes) ou de l'entreprise elle-même, ou sur leur traçabilité et leur conformité, en particulier lorsqu'elles sont intégrées dans un indicateur consolidé.

L'entreprise ou l'administration publique doit faciliter la fluidité des échanges de ses données, véritable patrimoine commun, entre les collaborateurs du groupe. Certaines entreprises ont mis en place une équipe transverse d'acteurs métiers légitimes pour décroquer les silos fonctionnels, promouvoir et faciliter les échanges de données entre équipes.

Dans certaines organisations, les métiers sont invités à se projeter et à réfléchir à **la valeur que les données produites à leur profit sont susceptibles d'apporter aux autres métiers de l'entreprise**, y compris en les croisant avec des données publiques. C'est donc une **vision de groupe ou même une vision systémique de l'écosystème qu'il convient de diffuser, et non pas une vision centrée sur le métier**.

Quelques entreprises ont fait le choix de systématiser le partage de la donnée, favorisant l'exploration de son utilisation, afin de stimuler l'émergence de nouveaux cas d'usage.

Retour d'expérience EDF : Acculturation et formation

Le déploiement de la stratégie Data au sein d'EDF nécessite à la fois d'acculturer le plus grand nombre de collaborateurs et de former certaines personnes ayant des besoins précis de montée en compétence dans le domaine de la data.

Le groupe a ainsi décidé de déployer un programme d'acculturation qui s'adresse à tout le monde, y compris aux dirigeants et managers. Ces derniers devant être des ambassadeurs de la transformation data. L'acculturation passe par de la communication de contenus permettant

⁵ Key Performance Indicator

d'expliquer les grands principes de cette transformation mais aussi de démystifier les principaux concepts de la data. C'est aussi l'occasion de faire connaître les cas d'usages data déjà réalisés et leurs bénéfices, tout en mettant en lumière les acteurs de la filière data qui ne sont pas toujours bien connus des opérationnels. Il est à noter que ce processus d'acculturation prend du temps et peut être boosté par les personnes et les réseaux déjà actifs dans la transformation numérique.

En ce qui concerne le processus de formation, il est pris en charge au sein d'une Académie du numérique qui a créé un catalogue de formation dédié à la data. Les populations cibles sont aussi bien des débutants que des acteurs de la filière data ayant des besoins de formation plus pointus. Les formations peuvent être assurées par des organismes externes ou bien par des acteurs internes, en l'occurrence des experts issus d'entités data de l'entreprise. Des parcours de formation sont spécialement prévus pour les personnes souhaitant se lancer dans une reconversion professionnelle. Cela répond ainsi à une problématique de *sourcing* interne des compétences data encore rares dans les entités. Ce catalogue de formation data est librement consultable par tous et la DSI⁶ Groupe en assure la mise à jour.

Sabine Macia, EDF

Former

Pour assurer l'adhésion à la stratégie data, l'entreprise doit être en mesure d'avoir une offre de **formation** ou **de mise à niveau** (*up-skilling*) **robuste**, tant pour les opérationnels que pour les managers, ou la direction. Plusieurs entreprises et administrations publiques font le choix de développer un programme pour acculturer l'ensemble des salariés (Dirigeants, Management, Terrain) et accélérer la transversalité et la fertilisation croisée de la filière valorisation de la donnée. Il faut cependant veiller à ce que la mise à niveau cible bien les profils intéressés et visés et donne le niveau espéré, de façon durable. En effet, une entreprise a mentionné avoir mis en place une formation fort bien notée par les participants en termes de contenu. Cependant, le résultat du test de connaissance, fait 3 mois plus tard auprès des mêmes participants, n'était pas à la hauteur des attentes concernant le niveau espéré. Cette formation a donc été complètement revue.

Pour les collaborateurs de la direction des systèmes d'information, la formation doit être plus poussée, notamment au plan technique (infrastructures d'acquisition des données, gestion, exposition, échange et valorisation, sauvegarde, archivage, purge) et doit inclure les règles de design et les modèles d'information métier (glossaire, mobilisation canonique des données). Apporter de la visibilité sur le périmètre des missions et les différentes compétences des équipes (direction data, pôle data de l'IT, métiers) permet aux collaborateurs de mieux appréhender les diverses parties prenantes et de comprendre l'articulation de la compétence de la DSI, des métiers, des experts externes.

Des entreprises ont fait le choix de cartographier les personnes travaillant dans le domaine de la data en interne pour identifier leurs compétences. Ainsi la communication est adaptée selon les fonctions ou les profils ciblés et les formations sont mieux ajustées.

Pour que la stratégie data puisse être mise en place avec succès, il faut se baser sur un cadre qui explique clairement comment les utilisateurs peuvent utiliser les données. Il est donc très important que les collaborateurs soient formés aux nouveaux outils et aux processus associés pour qu'ils se les

⁶ DSI = Direction des Systèmes d'Information

approprié et les utilisent correctement. Certaines entreprises et administrations publiques mettent en place des communautés data transverses pour favoriser cette acculturation et partager les bonnes pratiques.

Retour d'expérience Cnav : Acculturation à la qualité des données

La Cnav gère et est responsable de données pour ses propres besoins de régime général de retraite mais également pour des besoins émanant des autres régimes de retraite et/ou d'organismes de la sphère sociale (Ex : NIR – Numéro d'Inscription au Répertoire plus connu sous le nom de numéro de sécurité sociale).

La DSI de la Cnav a créé une direction de la gestion de la donnée qui porte notamment la gouvernance de la donnée. Cette direction propose une offre de service sur le *Master Data Management*, sur la gestion des échanges, sur la gestion de la qualité de la donnée et sur l'acquisition de la donnée pour l'ensemble de la DSI et des métiers consommateurs de la donnée. À ce titre, elle fait des sensibilisations et des formations sur ces différents thèmes. **L'ambition est d'aller vers une connaissance éclairée, juste et à forte valeur ajoutée des données dont la responsabilité est portée par la Cnav pour le compte de la sphère sociale.**

Bertrand Dubrulle, Cnav

Communiquer

Plusieurs entreprises et administrations publiques témoignent de l'importance de partager et de **communiquer** largement sur la stratégie data, sur sa **finalité et ses attendus** afin d'embarquer et de sensibiliser tous les salariés. Ce partage à l'échelle de l'entreprise facilite l'alignement entre ces différentes entités. Cela contribue à promouvoir une culture de la donnée. Le langage utilisé tient une place importante : éviter les termes trop techniques, et leur préférer des messages simples, plus clairs pour l'ensemble des parties prenantes afin de susciter l'adhésion de tous. **La stratégie data n'est pas qu'un sujet technique, elle relève de la transformation d'entreprise.**

Certaines données sont partagées au sein des organisations en *open data*, c'est-à-dire qu'elles sont disponibles en libre accès et peuvent être utilisées et partagées librement. Par définition, les données ouvertes doivent être accessibles, réutilisables et redistribuables sans restriction par n'importe quel utilisateur, ce qui nécessite une communication ad hoc pour une bonne utilisation.

La communication sur la stratégie data de l'entreprise ou de l'administration publique se fait également en externe auprès de son écosystème, mais aussi plus largement, à travers des hackathons⁷ ou des défis auprès des écoles, à travers des partenariats avec des startups, lors d'interventions dans les universités ou grandes écoles ou encore à l'occasion de stages. Les médias constituent aussi un bon moyen pour communiquer sur la stratégie data de l'entreprise, sur des nouveaux services ou sur la **disponibilité** de catalogues de données. Dans certaines entités, un ambassadeur de la data a été nommé dans l'équipe communication. Enfin des concours externes de type IT Night⁸, mettent également en avant des travaux sur la stratégie data auprès des médias.

⁷ Le hackathon désigne un rassemblement d'informaticiens réunis durant plusieurs jours et au minimum une nuit (souvent organisé sur un week-end) en vue de collaborer sur des sujets de programmation informatique pointus et innovants.

⁸ Lors de cette soirée, les professionnels de l'IT fêtent les plus belles innovations.

Appui de la Direction

Comme la stratégie data conduit à une transformation de l'entreprise, le sponsor de la Direction constitue un réel atout, qu'elle intervienne en tant que pilote ou en tant que partie prenante. Cela permet d'une part d'inclure l'ensemble des entités dans la construction de la stratégie data, dans l'étude des chaînes de valeurs essentielles au niveau du groupe puis dans la priorisation. D'autre part, ce soutien facilite la communication en interne et donc participe à la mise en place rapide de la valorisation des données. Enfin, ce sponsoring essentiel aplanit les difficultés rencontrées lors de l'élaboration de la gouvernance des données au niveau des équipes du SI, du digital et des métiers puis de sa mise en œuvre.

2.2.2 ÉTAPE 2 : DÉFINIR DES OBJECTIFS MÉTIERS PRÉCIS

Il est primordial pour l'entreprise ou l'organisation publique d'identifier les enjeux *business*, organisationnels, réglementaires et parfois financiers ainsi que les objectifs autour desquels construire sa stratégie data. Ces éléments vont lui permettre de déterminer les mesures, les indicateurs clés de performance (KPI) et les métriques pertinentes pour suivre l'impact de la stratégie data et piloter son déploiement. Cela aide également à identifier les sources de données pour la curation⁹ et l'analyse.

2.2.2.1 Enjeux business

Certaines entreprises ont construit leur stratégie data en réfléchissant aux cas d'usage qui répondent à leurs enjeux ou à leurs points de douleur, aux besoins des métiers, ou encore aux objectifs élargis de l'entreprise avec une approche systémique au niveau de leur écosystème (clients, fournisseurs, vendeurs et partenaires). D'autres se sont demandées en quoi la donnée pourrait contribuer aux objectifs stratégiques définis par la direction.

Retour d'expérience Hager Group : Leviers de la construction de la stratégie data

Une des ambitions de Hager Group est de devenir le conseiller numérique de confiance de ses clients en exploitant les systèmes qui gèrent et valorisent les données dans ses installations électriques et de gestion d'énergie. Sa stratégie d'apport de valeur vise à améliorer la gestion de l'énergie qui constitue un de ses axes de développement.

La data n'apporte plus seulement du support aux fonctions opérationnelles de l'entreprise, **elle génère à présent du *business* à mesure que ses systèmes deviennent connectés ou embarquent de l'intelligence artificielle dans ses produits.**

Sylvain Girard, Hager Group

⁹ La curation de contenu (étymologiquement du latin *curare* : prendre soin, et de l'anglais : content curation ou data curation) est une pratique qui consiste à sélectionner, éditer et partager les données et les contenus les plus pertinents au sein de l'écosystème pour une requête ou un sujet donné.

Un des participants indique qu'il s'est fondé sur l'analyse de la valeur *business* de ses cas d'usage data afin d'évaluer les bénéfices *business* qu'il pourrait tirer d'une stratégie data structurée.

Il est important de traduire la stratégie data par des petits pas qui apportent chacun de la valeur, autrement dit en *quick wins*.

Une autre approche partagée par plusieurs participants du groupe de travail « Gouvernance et architecture data » consiste à identifier les données au cœur des usages et des initiatives puis à évaluer le délai de récupération d'un coût d'investissement (*pay back*) et la contribution de ces données aux enjeux stratégiques et aux priorités, au regard des programmes clients et des nouveaux services.

Une démarche *bottom up* est également, parfois utilisée : elle consiste à identifier les investissements data déjà réalisés afin de les coordonner au niveau de l'entreprise ou de l'administration publique. Dans ce cas, les participants préconisent de mener en parallèle les démarches *bottom up* et *top down* avec des points de rencontre afin que la stratégie et le terrain restent connectés.

Enfin certaines organisations **partent de** leur raison d'être pour élaborer leur stratégie data.

Retour d'expérience du groupe SAUR : Exemple de déclencheur de la stratégie data

Le cœur de métier de SAUR est de gérer l'eau. La tension sur la ressource en eau s'accélère. Toutes les régions sont touchées en France, ce qui n'était pas le cas auparavant. SAUR est parti de ses enjeux *business* mais également environnementaux et socio-démographiques pour déterminer sa stratégie data.

L'ambition de SAUR est de défendre l'eau grâce au numérique avec la mise en place des installations nécessaires pour télésurveiller l'ensemble du réseau, traiter les données remontées des usines, des équipements et du réseau. C'est pourquoi le numérique du groupe SAUR s'est organisé pour mettre la donnée au cœur du système d'information. La donnée circule entre les usagers, les clients, les territoires, les équipements les installations et les différents systèmes d'information de l'eau.

Alice Guéhennec, Groupe SAUR

2.2.2.2 Enjeux organisationnels

La stratégie data est construite dans certains cas pour répondre à des enjeux de type organisationnels. Ainsi la stratégie data est au service des métiers pour leur permettre d'accomplir leur mission plus facilement.

Retour d'expérience de Pôle Emploi : Exemple de déclencheur de la stratégie data

Le déclenchement de la stratégie Data de Pôle Emploi résulte de la fusion entre l'ANPE (en charge du retour à l'emploi des demandeurs) et les ASSEDIC (en charge du versement des indemnités aux demandeurs d'emploi). Pôle Emploi cherchait à procurer une autonomie aux demandeurs d'emploi familiaux avec l'utilisation du numérique, pour que les conseillers se

focalisent sur la personnalisation de l'accompagnement des demandeurs d'emploi moins aguerris. Le numérique avec l'usage des données, offre cette possibilité.

Patrick Mahu, Pôle Emploi

Les entreprises ou administrations publiques concernées cherchent à inclure des données dans leurs prises de décisions, autrement dit à mettre en place une organisation « *data driven* ». Une fois les objectifs annuels métiers mesurables avec des indicateurs et des données clairement définis, elles s'organisent pour que les métiers pilotent davantage les initiatives par la valeur apportée.

2.2.2.3 Autres enjeux

Plusieurs entreprises ont évalué toute la chaîne de valeur de leurs produits, solutions et services de bout en bout afin de déterminer les autres bénéfiques qu'il y aurait à disposer d'une stratégie data qui assure des données disponibles et de qualité et qui intègre des modules de traitements de la donnée, de sa production jusqu'à son utilisation en interne, comme au sein de son écosystème externe.

Dans certains cas, les entreprises et organisations publiques ont besoin de traçabilité et de sécurité sur leurs services et produits. Cela nécessite de pouvoir auditer les accès aux données. Ce type d'audit est particulièrement requis lorsque la donnée est partagée avec des entités externes à l'entreprise. Des règles claires d'APIsation, de catalogage et de circulation des données doivent alors être mises en place au niveau de l'entreprise ou de l'administration publique.

Certains enjeux réglementaires nécessitent que des données dites sensibles soient particulièrement identifiées (ex : données personnelles, confidentielles défense, secret médical, etc...), afin de pouvoir gérer certaines contraintes liées à leur stockage, leur conservation, leur accès, leur partage.

Enfin, il peut s'agir d'enjeux financiers, pour lesquels l'exploitation de la data permet des optimisations ou des estimations (fiscalité, investissements, coûts...).

2.2.3 ÉTAPE 3 : IDENTIFIER LES DONNÉES PERTINENTES À UTILISER

La stratégie data nécessite de capturer les bonnes données, après les avoir identifiées, puis de partager leurs définitions dans un catalogue de données, pour atteindre les objectifs fixés tout en intégrant des principes de sobriété. En effet, les entreprises et administrations publiques collectent des données, mais leurs sources doivent être « nettoyées » et « certifiées » afin que les usagers utilisent des données adaptées aux différents cas d'utilisation et aux objectifs tout en ayant la certitude que celles-ci proviennent de sources fiables et précises conformes aux normes en place. D'où l'importance de prendre en compte les préoccupations des services juridiques, financiers, RH, IT et conformité. C'est pourquoi les entreprises et organisations publiques veillent à la connaissance du patrimoine de données et à leur qualité. Une fois la donnée récupérée, il faut la rendre fiable, puis l'intégrer au dictionnaire de données. Des tableaux de bord sont mis en place pour mesurer la qualité de la donnée même si celle-ci n'est pas toujours un prérequis. En revanche, il semble primordial de connaître la sensibilité de la donnée pour déterminer quel niveau de contrôle de sa qualité mettre en œuvre et maîtriser de façon avérée l'évaluation de son degré de qualité. Les entreprises et administrations publiques mettent en place des référentiels de données d'entreprise et un « Master Data

Management » ou MDM, en français « Gestion des données de référence », qui constitue un élément central de la stratégie data. Le MDM utilise un ensemble d'outils permettant de stocker, gérer et diffuser les données de référence au sein d'une organisation. Il permet de s'assurer que celles-ci sont correctement identifiées, utilisables sans risque, exemptes d'erreur et de bonne qualité.

Afin de mieux piloter la stratégie data en mesurant l'avancée de son déploiement et ses bénéfices, les entreprises et administrations publiques mettent en œuvre des métriques et KPI.

2.2.4 ÉTAPE 4 : PRÉPARER L'ARCHITECTURE ET LES BRIQUES TECHNIQUES

Un autre prérequis pour faciliter la mise en œuvre de la stratégie data est de définir une architecture data cible, atteignable en plusieurs étapes intermédiaires. L'architecture mise en place pour les données doit être compatible avec les standards d'architecture du SI relevant du transport de données (messages, fichiers) ou de l'accès aux données (API ou Web Services), de sorte à ne pas impacter les systèmes sources de ces données. Autrement dit, l'architecture doit faciliter l'accès aux données. Le plus compliqué consiste à sécuriser l'information ou la donnée de manière simple et efficace tout en garantissant le partage de celle-ci. Ensuite, il faut préparer l'ensemble des briques fonctionnelles nécessaires pour mettre en œuvre la stratégie data. Par exemple, pour développer la vente en ligne, il est important de savoir gérer les stocks en temps réel.

Les entreprises et administrations publiques se connectent aux sources de leurs données et prennent en compte toutes les façons dont ces dernières peuvent être enrichies. Cela impose parfois d'abandonner l'entrepôt de données traditionnel, d'un seul tenant, pour adopter une approche multiple qui permet la manipulation de grandes quantités de données, rapidement et de façon agile.

3 MISE EN OEUVRE DE LA STRATÉGIE DATA

Une fois la stratégie data de l'entreprise déterminée, il reste à l'implémenter que ce soit au niveau des fondations techniques, de la gouvernance, de l'organisation ou des processus. Il faut également pouvoir piloter sa mise en œuvre et en suivre les bénéfices avec des mesures et des indicateurs bien identifiés.

3.1 FONDATIONS TECHNIQUES

Les participants au groupe de travail « Gouvernance et architecture data » notent des principes communs dans les fondations de l'architecture des entreprises, pour mettre en place leur stratégie des données, qui s'articulent autour de plusieurs types de briques : **la captation et le stockage des données**, leur **gouvernance**, leur **exploration, exposition et exploitation** (dont le calcul et l'analytique) incluant des briques de *business Intelligence*, et enfin **la stratégie logicielle et de l'infrastructure**.

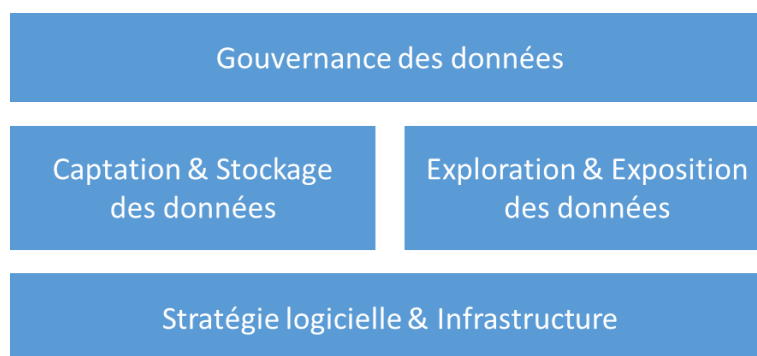


Figure 2 Briques communes dans les fondations de l'architecture pour déployer la stratégie data

La mise en qualité des données est une étape incontournable avant leur stockage.

Les choix technologiques sont variés lors de la mise en œuvre de la stratégie data des entreprises. Pour chacune des briques de l'architecture data, les entreprises et les administrations publiques utilisent les solutions les mieux adaptées, ce qui multiplie les technologies et solutions utilisées, engendre des problèmes en matière de coûts de possession, de dette technologique, et d'entretien des compétences sur un panorama logiciel complexifié. C'est pourquoi plusieurs d'entre elles cherchent maintenant à trouver un bon compromis entre des fonctionnalités très riches, l'usage des nouvelles technologies et l'interopérabilité entre les solutions afin d'avoir une solution globale.

Les entreprises et administrations publiques analysent le risque sur le choix des outils et cherchent à diminuer leur dépendance à ces solutions technologiques, et aux éditeurs associés. La capacité de changer rapidement d'éditeur est recherchée, ainsi que la possibilité d'exporter tout le contenu. D'une part, les entreprises manquent de visibilité sur l'évolution des coûts des outils dans le temps et d'autre part, les *business models* des éditeurs peuvent être amenés à évoluer très vite. Des solutions en open source peuvent être privilégiées pour ne pas être captifs. Il s'agit de « choisir ses dépendances ».

Enfin plusieurs entreprises se penchent sur le *data mesh* qui propose un ensemble de pratiques et de principes d'organisation pour appréhender les nouveaux défis posés par la gestion de la donnée à

grande échelle. En effet, le *data mesh* repose sur une gestion et une exposition, décentralisées de la donnée pensée en tant que produit, et assurant les fondamentaux suivants :

- Donner de l'autonomie en permettant une gestion locale des données ;
- Faire grandir le système en y ajoutant facilement des nœuds de manière décentralisée ;
- Conserver une cohérence forte de la plateforme de données (technique et organisationnelle).

3.1.1 CAPTATION & STOCKAGE DES DONNÉES

Les entreprises et administrations publiques collectent les données via des capteurs (compteurs, sondes, IHM de saisie, ...) et des logiciels d'alimentation de ces dispositifs de stockage (ELT, ETL, messaging, scripts, ...) et utilisent comme dispositifs de stockage un *data lake* (lac de données) et des *data warehouse* (référentiels de données structurées). Les *data hubs* leur servent de médiateurs et de points de partage des données. Les *data warehouses*, les *data lakes* et les *data hubs* ne se substituent pas les uns aux autres, ils sont complémentaires et, combinés, ils permettent la mise en œuvre de la stratégie data de l'entreprise.

3.1.1.1 Data Lake

Un *data lake* est un espace unique regroupant toutes les données provenant de tous les systèmes de l'entreprise, voire de l'extérieur de l'entreprise. Il contient un ensemble maîtrisé et qualifié de données opérationnelles, c'est-à-dire les **données brutes**, ingérées et stockées dans leur format d'origine (des systèmes sources), par exemple, à des fins de traçabilité et de reconstitution éventuelle en cas de problème. Il n'y a pas de hiérarchie ou d'organisation entre les différents éléments de données. Des modèles de données et des catalogues sont nécessaires à l'exploitation et à la bonne organisation des *data lakes*.

Le *data lake* peut être interrogé pour obtenir des données pertinentes, et ce plus petit ensemble de données peut ensuite être analysé pour aider à répondre à des problématiques d'entreprise.

Les *data lakes* sont, en général, une base solide pour la préparation des données, les rapports, la visualisation, l'analyse approfondie, la *data science* et le « *machine learning* ».

Un participant du groupe de travail « Gouvernance et architecture data », dont l'organisation regroupe en interne de nombreux métiers, ce qui se traduit par des disparités en termes de données et de traitement de celles-ci, a choisi d'aller vers un *data lake* centralisé au niveau institutionnel de son organisation, pour l'ensemble de ses entités. Ce choix d'un *data lake* centralisé permet une recentralisation du champ des données ce qui facilite le maillage des données, source d'apport de valeur.

3.1.1.2 Data warehouse

Le *data warehouse* est un dépôt central de **données structurées, modélisées et filtrées** qui ont déjà été transformées dans un but spécifique, autrement dit, c'est une base de données relationnelle hébergée sur un serveur dans un data center ou dans le cloud. L'entreprise ou l'administration publique doit choisir tout d'abord, si elle souhaite garder ses ressources en interne, ou en externaliser

une partie ou l'intégralité. Cela dépasse la seule stratégie data et doit être décidé dans la stratégie de l'entreprise.

Le *data warehouse* recueille des données de sources variées et hétérogènes dans le but principal de soutenir l'analyse et de faciliter le processus de prise de décision. L'alimentation du *data warehouse* peut être basée sur le processus ETL (Extract, Transform, Load, mais aussi sur d'autres processus comme par exemple le *messaging*, le moteur de recherche) permettant de charger les données issues des différentes applications du SI opérationnel.

Afin de le rendre plus robuste, des entreprises et administrations publiques ont fait le choix d'appuyer leurs *data lake* et *data warehouse* sur plusieurs instances dans différents sites géographiques de data center et d'autres utilisent des technologies différentes pour opérer leur *data warehouse*.

3.1.1.3 Data Hub ou hub de données

Le hub de données est un modèle d'architecture visant à favoriser la collecte, le maillage et la valorisation des données, issues de multiples sources, à des fins de mise à disposition de cas d'usage métier opérationnels. Le but du *data hub* est de faire converger les données vers un même référentiel, indépendamment de leur provenance, et de leur métier (c'est ce qu'on appelle le processus d'uniformisation des données). Les données qui transitent dans cette plaque tournante qu'est le *data hub* sont améliorées : cataloguées, indexées, taguées, sécurisées et harmonisées. D'une part, la performance technique est assurée pour les utiliser au mieux, d'autre part, la valeur *business* est renforcée. Le *data hub* ne propose que des services de fourniture de données qu'il a obtenus par ailleurs (via *dataset*/fichiers ou *open data* / API).

Le silotage inhérent à l'activité de l'entreprise à cause des différents processus métier (le processus d'achat, le processus de vente, le processus d'exploitation, le processus de ressources humaines et bien d'autres) rend cette tâche de transversalité très compliquée. D'où l'importance du *data hub*.

Quelques entreprises font le choix de centraliser le stockage des données pour favoriser le désilotage de celles-ci et pour certaines, elles s'orientent alors pour cette brique technologique soit vers des solutions cloud qui est un moyen d'assurer la centralisation soit vers des solutions sur site (*on premise*).

3.1.2 BRIQUES DE DATA GOUVERNANCE

La gouvernance des données est l'ensemble des processus, compétences, ressources, pratiques, méthodes et outils, mis en œuvre pour favoriser l'efficacité des usages métier et optimiser l'exploitation et la qualité des données, au service des enjeux métier. Les entreprises et les administrations publiques sont unanimes pour souligner l'importance de s'outiller pour mettre en place une gouvernance des données commune au niveau du groupe. Partir d'une approche globale permet de favoriser le choix de solutions, utilisées au niveau de la région ou du produit, compatibles entre elles. La stratégie data étant transversale à toute l'entreprise par essence, il importe de **veiller à une certaine cohérence dans le traitement de la donnée au niveau global**. De même la cohérence dans le stockage et le traitement des données au niveau de l'entreprise, favorise leur utilisation dans des offres et services ou contribue à leur amélioration (pertinence du service, identification des candidats à la fraude...).

Les règles de gouvernance doivent être précisément établies. Cependant, comme elles sont sujettes à des évolutions pour différentes raisons (ex : une nouvelle organisation, une acquisition, etc.), les organisations cherchent des solutions de gestion de données flexibles.

Les outils et les principes d'architecture sont choisis en fonction de la gouvernance que l'on souhaite mettre en place.

Les systèmes d'informations sont de plus en plus hétérogènes (suite à des fusions et acquisitions, à des choix de solutions ou de technologies variés, etc.) et les données proviennent de sources de plus en plus nombreuses. Les données sont cartographiées en fonction de leur nature, de leurs usages, de leur localisation (production et stockage). Définir des principes et des règles permet d'éviter les doublons, et d'identifier l'application source.

La gouvernance des données nécessite notamment l'utilisation d'outils tels que les catalogues de données et les dictionnaires de données.

- Le **dictionnaire de données** - *data dictionary* - regroupe les informations des bases de données et des systèmes informatiques de l'entreprise ou de l'administration publique. Il regroupe également les définitions et les descriptions détaillées des jeux de données (regroupés dans des tableaux) et de leurs valeurs (champs). Le dictionnaire des données s'adresse en priorité aux équipes IT, mais est de plus en plus utilisé par les métiers, dans le cadre de la démocratisation de l'accès à la donnée, sous-tendu dans la stratégie data. Il permet de standardiser les informations sur ses données : chaque source de donnée est décrite dans un dictionnaire de données selon le standard choisi : convention de nommage, source, typologie de donnée, taille, format, description, exemple... Il est le point d'entrée unique vers toutes les informations des différentes sources de données.
- Le **catalogue de données** - *data catalog* - est un logiciel informatique (développé en interne ou acheté) permettant d'avoir une vue unifiée de l'ensemble des informations servant à décrire les données dans l'entreprise (type de fichier, origine, date, etc.). Il expose le descriptif (métadonnées) du patrimoine des données hébergées dans le SI et permet de centraliser l'information et la standardiser. Il permet de gérer les droits d'accès aux bases de données, de savoir qui est « *Owner* » (propriétaire) ou « *Steward* » (agent ou responsable) des données. Cet outil est donc très utile aussi pour responsabiliser les acteurs de la donnée dans les organisations et savoir à qui s'adresser (désigner par exemple qui est censé mettre à jour les informations). Selon les traductions locales et le niveau d'information, les champs ne sont pas toujours utilisés correctement ce qui est cause de préjudice sur le traitement. Des participants préconisent de veiller aux objectifs de chaque champ pour assurer un catalogue de données de qualité.

Le catalogue de données s'adresse plutôt aux usagers orientés métiers et aux maîtres d'ouvrage. Son déploiement se fait progressivement après parfois des négociations ardues avec la direction lorsque la stratégie data est *bottom up*.

Plusieurs entreprises ont mis en place la gouvernance de la stratégie data en commençant par la Gestion des Données de Référence ou *Master Data Management*, mais elles constatent qu'elles sont perdues par la complexité. Selon ces dernières, il faut d'abord homogénéiser les outils utilisés car la culture et la maturité sont à des stades très différents au sein d'un même écosystème. La gouvernance peut s'appuyer sur une analyse sophistiquée. Cependant, il faut éviter de tomber dans l'engrenage de solutions toujours plus coûteuses. En outre, il faut ajouter à cela qu'une transformation de ce type est longue en raison de sa dimension humaine.

3.1.3 EXPLORATION, EXPLOITATION ET BUSINESS INTELLIGENCE

Les participants soulignent l'importance de procurer un maximum d'autonomie aux utilisateurs à travers deux modes : le mode gouverné et le mode self-service, avec des règles et principes clairement définis et partagés. Cependant, partager une définition commune de la donnée au sein de l'entreprise n'est parfois pas aussi simple qu'il y paraît. Une même dénomination peut pointer sur deux données différentes ou à l'inverse deux dénominations différentes correspondent à une même donnée.

Il y a également un enjeu sur la qualité des données, en raison de processus manuels. Les données sont parfois enrichies par les métiers d'où l'importance de sensibiliser les collaborateurs à la qualité de la donnée.

Avec l'arrivée du cloud et de ses solutions SaaS, les métiers sont tentés d'acquérir de nouveaux outils. Face à un effectif croissant d'outils, plusieurs entreprises tentent de rationaliser l'ensemble des solutions. L'arrivée sur le marché de jeunes diplômés formés aux outils récents (R, Python, etc.) et leur embauche dans les entreprises et administrations viennent aussi challenger les outils déjà en place.

La Business Intelligence (BI) désigne les technologies, applications et pratiques de collecte, d'intégration, d'analyse et de présentation de l'information. L'objectif de la BI est d'analyser les données afin de soutenir une meilleure prise de décision dans les domaines métiers : commercial, marketing, finance, RH, etc.

Lorsque les entrepôts de données BI (*data warehouse*) sont vétustes, la priorité est souvent de moderniser toute la plateforme data avec l'actualisation des briques de BI. Plusieurs entreprises cherchent à développer la culture de la *Self BI* avec, au préalable, validation du droit d'accès à la donnée.

Certaines entreprises utilisent des outils de virtualisation de données. Ces outils permettent d'accéder aux données, avec leur contexte grâce à des APIs, tout en masquant les aspects techniques tels que l'emplacement, la structure ou la langue d'accès. Cependant, ces organisations sont attentives à rationaliser le nombre d'outils de visualisation, afin d'avoir des données intégrées dans une interface unique qui facilite l'aide à la décision.

À noter aussi l'accélération des besoins en données, dans le cadre de la montée en puissance des solutions d'intelligence artificielle (*machine learning, deep learning*). Ces nouveaux besoins viennent renforcer l'intérêt et la nécessité d'une stratégie data solide.

3.1.4 STRATÉGIE LOGICIELLE ET INFRASTRUCTURE

Une même entreprise a souvent de multiples occurrences de référentiels de données, localisés à de nombreux endroits et dont le nombre s'accroît avec les acquisitions. Cela crée de nouveaux défis pour l'accès aux données, en particulier dans une grande organisation. Il est clair que la technologie seule ne résout pas les problèmes liés aux données. D'où la question pour certaines organisations de concevoir une architecture fonctionnelle pour des données unifiées aussi appelée UDA (*Unified Data Architecture*). Ce problème n'est pas simple à résoudre et les entreprises peuvent adopter plusieurs approches pour y répondre, chacune ayant ses propres forces et faiblesses. Il n'y a pas de solution unique.

La notion de sécurité pour les solutions cloud est plus ou moins bien définie au niveau des entreprises et administrations publiques. Elle s'articule sur plusieurs axes : la confidentialité (chiffrement des

données directement sur les unités de stockage avec des clefs sur site), l'intégrité (non dégradation de la donnée lors de ses manipulations et de son cycle de vie, et sa transformation maîtrisée), et la disponibilité (résilience en cas de rupture d'accès physique).

Des entreprises soulignent le gain réel apporté par les services cloud mais soulignent aussi le problème de dépendance face aux éditeurs. On note plusieurs écoles non exclusives qui se concrétisent par des modèles hybrides pour stocker les données :

1. Stockage sur le site de l'entreprise ou de l'administration publique (on-premise). Ce choix des entreprises de garder leurs données sur des data centers sur site peut être motivé par une politique qui considère que les activités d'exploitation doivent rester en interne ou par des besoins de confidentialité, de sécurité et de conformité au RGPD ou pour des raisons d'immunité aux lois extraterritoriales. En effet, les entreprises et organisation publiques s'accordent à dire qu'en moyenne, entre 10 et 35% de leurs données, avec une grande variabilité en fonction de leur secteur d'activité, **ne peuvent pas migrer sur le cloud public qui ne répond pas à leurs besoins en termes de confiance**. Elles veulent en effet **garantir la sécurité de leurs données et traitements les plus stratégiques et attendent des offres industrielles adaptées à ce besoin**¹⁰. Le référentiel « [Cloud de confiance](#)¹¹ » élaboré par le Cigref traduit en exigences fonctionnelles et objectivables la confiance telle que l'expriment les adhérents du Cigref en matière de solutions et services cloud. En l'absence de solution industrielle répondant à ce besoin, certaines entreprises ou administrations publiques construisent actuellement leur cloud interne.
2. Stockage sur un cloud public. La flexibilité offerte par le cloud est un avantage mentionné à plusieurs reprises, de même que la garantie de sécurité qu'il apporte, sécurité que les entreprises n'arrivent pas à gérer avec le même degré d'exigence en interne. Des solutions hybrides cloud privé/public et semi privé sont également choisies par certains. Plusieurs entreprises, actuellement en cours de migration, commencent leur trajectoire vers des solutions SaaS et PaaS. Cette migration progressive permet de définir des règles de gouvernance en fonction de la nature des données, du type de flux, du lieu d'hébergement, etc. Des organisations participantes font état d'écosystèmes data vétustes, composés de plateformes différentes, avec des technologies diverses. L'une d'elles, dont environ 70 à 80% des données sont dans l'ERP, utilise des plateformes historiques solides, bien que décentralisées et hétérogènes. Cependant, elle a choisi un cloud public américain pour répondre à ses enjeux présents et futurs. D'autres ont choisi des solutions multi-cloud afin de répondre à leurs besoins en termes de services¹².

3.2 GOUVERNANCE

La gouvernance, prérequis pour bien décliner la stratégie data, répond à plusieurs enjeux : connaître les données, maîtriser leur qualité, harmoniser les données de référence, assurer leur protection et leur confidentialité, gérer les accès et enfin acculturer les équipes à la donnée. Il ressort des échanges au sein du groupe de travail « Gouvernance et architecture de la data » l'importance d'articuler la

¹⁰<https://www.cigref.fr/position-des-associations-utilisateurs-allemande-et-francaise-sur-le-projet-de-schema-europeen-de-certification-pour-le-cloud>

¹¹ <https://www.cigref.fr/cigref-publishes-its-second-version-of-the-trusted-cloud-reference-document>

¹² Pour aller plus loin sur le sujet, nous vous invitons à consulter le rapport Cigref « [Stratégies de migration dans le cloud. Une aventure stratégique pour l'entreprise](#) », novembre 2021. Des travaux complétant ce livrable sont en cours et seront publiés fin 2022.

gouvernance de la stratégie data avec celles des autres stratégies mises en place au sein de l'entreprise ou de l'administration publique comme la stratégie des SI, la stratégie des métiers et la transformation digitale, qui concourent toutes à la stratégie de l'entreprise. De plus, la gouvernance de la stratégie data doit pouvoir s'articuler au niveau du monde, d'une région du monde, du marché.

La gouvernance valide l'alignement du *business* avec la stratégie data et facilite la priorisation des services à implémenter. Au regard des enjeux portés par la gouvernance, le comité de gouvernance doit avoir du poids à l'échelle de l'entreprise. Plus généralement, plusieurs entreprises soulignent l'importance de mobiliser des sponsors dans le comité exécutif et pour cela de trouver des arguments pour convaincre ces derniers d'appuyer l'instauration de la gouvernance de la data dans l'entreprise. Selon les entreprises, la gouvernance data est prise en charge par le CIO (*Chief Information Officer*), par le CDO (*Chief Data Officer*) ou encore par les métiers.

La conformité et la qualité des données imposent à la fois une exigence de rigueur, d'intégrité et de compréhension de la sémantique des données. La problématique de la donnée et de la sémantique est d'autant plus importante aujourd'hui à l'âge de la digitalisation des entreprises que beaucoup de systèmes différents sont en place. On urbanise en permanence et de plus en plus de nouveaux systèmes. Les problèmes liés à la sémantique ont toujours été présents. Mais, si on les traite au fil de l'eau, ce qui n'est plus le cas aujourd'hui, on se laisse submerger. Auparavant, on posait les normes puis on interconnectait les systèmes. Aujourd'hui on présente une procédure totalement différente : tout d'abord mise en place de l'interconnexion, puis ensuite mise en place des normes.

La normalisation des règles de nommage permet d'unifier les facettes utilisées dans les différents pays ou même dans différentes régions d'un même pays. La normalisation du glossaire métier est aussi un point essentiel. Une même donnée qui n'aurait pas la même définition entre distribution et opération, serait un véritable problème dans le quotidien de l'entreprise (ex : longueur, largeur, hauteur, il faut une définition commune pour vendre ses produits, remplir les camions qui les acheminent etc.). Il y a donc un impératif de normalisation des données dans le SI.

3.2.1 CONNAÎTRE LES DONNÉES (PROCESSUS « KYD KNOW YOUR DATA »)

Les participants soulignent l'importance de passer de « l'idée de données utiles », vue comme un concept, à une réalité concrète. C'est pourquoi les entreprises et administrations publiques réfléchissent à la façon de valoriser réellement le partage de la donnée entre équipes en faisant émerger des cas d'usages. Une des priorités est donc d'identifier les données qui génèrent un gain *business* ou un gain sur les processus en calculant le ROI de la donnée, tout en identifiant les risques. Pour cela, les collaborateurs de l'entreprise ou de l'écosystème de l'entreprise doivent savoir où les données se trouvent avec l'assurance de leur mise à jour.

C'est pourquoi, les participants du groupe de travail « Gouvernance et architecture data » préconisent de déterminer le leader de telle ou telle partie de la donnée, et d'identifier celui qui gère le partage de la donnée. La mise en place d'une gouvernance de la donnée au niveau de l'entreprise avec un outillage adapté, permet tout au long du cycle de vie de la donnée de pointer les responsabilités lors des différentes actions à mener. En effet, la gouvernance répond à plusieurs objectifs : assurer la qualité de la donnée tout au long de son cycle de vie, faciliter son utilisation en offrant une documentation actualisée du patrimoine des données et de la manière de l'utiliser, suivre l'évolution de sa consommation, s'assurer de l'adéquation avec les besoins des consommateurs et enfin s'assurer de son bon effacement en fin de vie. Pour partager et utiliser la donnée, il faut mettre en place des

ontologies et des règles communes au sein de l'écosystème de l'entreprise. Une ontologie est un ensemble structuré de concepts, permettant de donner un sens aux informations, et qui a pour premier objectif de modéliser un ensemble de connaissances dans un domaine donné. Les directions métiers qui produisent la donnée ont le devoir d'en faciliter le partage, afin de favoriser la création de valeur pour l'entreprise.

3.2.2 MAÎTRISER LA QUALITÉ DES DONNÉES

Pour utiliser les données, les collaborateurs doivent avoir l'assurance de leur qualité. Chaque donnée doit être bien cataloguée, identifiée, définie (plus particulièrement pour les indicateurs), à jour et exploitable. Il est important de savoir mesurer la qualité de la donnée pour l'améliorer jusqu'à ce que la valeur de référence soit atteinte. Cette analyse est du ressort des métiers car la qualité répond à leurs besoins. Cette mesure de la qualité peut souvent s'appuyer sur la Gestion des Données de Référence (MDM), qui permet de vérifier que la valeur d'une donnée de type nomenclature est bien déclarée dans son référentiel associé. La qualité et la maturité de la donnée, évaluées par les audits internes, ainsi que son flux doivent être contrôlés. Ce niveau d'exigence permet de maximiser l'apport des données dans l'exécution des missions. Ainsi les métiers, les data analysts et les *data scientists* peuvent l'exploiter en toute autonomie et de façon pertinente.

Plusieurs participants veillent au sein de leur organisation à ce que le producteur de la donnée garde bien en compte que la **donnée est au service des usages**, lorsqu'ils définissent les règles de gouvernance. **Les règles visent donc bien les consommateurs potentiels de la donnée et doivent prendre en considération la finalité de la donnée.**

3.2.3 CONFORMITÉ RÉGLEMENTAIRE & PROTECTION

Il est nécessaire d'identifier les données sensibles ou « à valeur » afin d'adapter l'endroit où les stocker : en interne, chez des partenaires ou dans le cloud. L'accès aux données sera ainsi déterminé en fonction de leur sensibilité. Le responsable de la donnée, souvent le DPO¹³ ou ses instances, intègre les contraintes réglementaires associées et les partage aux utilisateurs de ces données sensibles. Certaines données confidentielles nécessitent le support d'un expert en sécurité afin de sensibiliser les collaborateurs aux enjeux de sécurité. L'objectif est de maximiser l'usage de la donnée avec des traitements analytiques tout en minimisant les risques de non-conformité, ou ceux liés à l'éthique, provenant de ces usages.

3.3 ORGANISATION

Lors du déploiement de la stratégie data, plusieurs participants préconisent de définir clairement les missions des différents acteurs de la data au sein de l'entreprise ou de l'administration publique (Métiers, IT et digital), et de bien articuler ces missions entre elles. Cependant, on note une certaine disparité dans le descriptif des dénominations des différents métiers de la data et de leurs missions. Le groupe de travail Cigref « [Nomenclature des profils métiers du SI](#) » actualisera en 2023 les métiers

¹³ DPO = Data Protection Officer, ou « Délégué à la protection des données »

liés à la data qui évoluent rapidement¹⁴. Certains groupes ont nommé un responsable des données¹⁵ (*Chief Data Officer*) dans chaque division métier afin qu'il décline la politique de la donnée en son sein.

L'organisation doit assurer la mise en cohérence du patrimoine des données qui constitue un vrai enjeu de stratégie des données.

L'organisation globale vise un dispositif communautaire et collaboratif. L'objectif est de s'inscrire dans la démarche agile de l'entreprise, en respectant les fonctions des métiers et de la DSI, et de les faire collaborer dans l'évolution du dispositif.

Lionel Naessans, IMSA

Aujourd'hui, les entreprises et administrations publiques participant au groupe de travail « Architecture et gouvernance data » ont une organisation pour mettre en œuvre la stratégie data plutôt centralisée mais elles visent la décentralisation dès lors qu'elles auront acquis davantage de maturité. Même si la donnée prend une place importante dans les entreprises, l'IT prend, en général, le lead en attendant que les différents métiers se l'approprient. Le caractère centralisé souligne que les entreprises sont très focalisées sur le caractère big data.

3.3.1 FÉDÉRATION PLUTÔT QUE DÉCENTRALISATION

Si on reste sur un modèle centralisé, on risque de limiter l'empreinte de la gouvernance des données, ainsi que l'appropriation des bonnes pratiques par tous les acteurs mobilisés sur le SI. La décentralisation permet de démultiplier les actions. Il faut trouver un équilibre entre ce qu'on centralise et ce qu'on décentralise. Plutôt que de parler de décentralisation qui induit le principe de subsidiarité, principe selon lequel une autorité centrale ne peut effectuer que les tâches qui ne peuvent pas être réalisées à l'échelon inférieur, certains participants préfèrent parler de "fédération". La fédération, en effet, doit permettre à chacun d'accéder à la data même si elle est dans des silos au sein de l'entreprise. Ces entreprises cherchent à fédérer les parties prenantes de la donnée au sein d'un réseau tout en conservant la responsabilité des métiers dans la gestion de celle-ci selon leur périmètre. Ainsi toutes les parties prenantes ont accès à la donnée. Pour faciliter la fédération dans certaines entreprises, l'IT propose des programmes d'outillage transverses aux métiers avec des centres d'expertise.

3.3.2 CONTRÔLE ET RESPONSABILITÉ DE LA DONNÉE

Dans plusieurs entreprises et organisations publiques, les métiers exercent un contrôle plus fort qu'avant sur la donnée et expriment une volonté de prendre en main la gouvernance data. En effet, la maîtrise de la gouvernance leur assure la qualité de la donnée requise afin de créer de la valeur. C'est

¹⁴ Pour rappel, l'apparition d'un profil métier générique est l'aboutissement d'un processus qui prend environ deux à trois ans : au départ des compétences spécifiques, de manière disparate, apparaissent dans les profils existants, puis un consensus se fait sur la définition de ces compétences. Quand une maturité est atteinte sur ces compétences et nécessitent une spécialisation, des profils spécifiques apparaissent, dont la description fait progressivement consensus. Ils peuvent alors être intégrés dans la Nomenclature RH.

¹⁵ Voir la description de sa mission dans la nomenclature des profils métiers du SI du Cigref dans l'espace Nomenclature des profils SI du Cigref, <https://www.cigref.fr/nomenclature-rh> :

la raison pour laquelle les organisations cibles à atteindre, concernant la gouvernance, donnent la main aux métiers. Plusieurs participants estiment d'ailleurs que cette prise en main par les métiers qui vient naturellement avec la maturité data, est une évolution souhaitable.

Cependant, certaines données sont transversales à l'entreprise et dans ce cas se pose la question du propriétaire de la donnée. En prendre la responsabilité induit un coût, des ressources et des compétences pour la gérer. Plusieurs entreprises ont organisé dans ce cas un modèle de responsabilité sur les populations de données.

En règle générale, les participants s'accordent pour dire que la donnée doit toujours être rendue disponible, à moins qu'il existe un risque. Dans ce cas, le responsable de la donnée, généralement le *Data Protection Officer (DPO)*, doit l'exposer au management. Si le risque est grand et que la donnée n'a pas beaucoup de valeur, elle n'est pas transmise car la mise en balance du bénéfice/risque est faible. Mais dans le cas où la donnée a une grande valeur et que sa publication engendre un grand risque, il appartient au management de trancher.

3.3.3 PROCESSUS

Plusieurs défis sont posés concernant l'identification et la qualification des cas d'usage dans la gestion du portefeuille et la priorisation de ces derniers. Comment prioriser les cas d'usage suite à une première évaluation par un PoC¹⁶ ou MVP¹⁷ : par la capacité d'un passage à l'échelle, par la valeur, par un *business case* solide ? Des cas d'usage transverses à plusieurs domaines sont-ils susceptibles d'apporter plus de valeur ?

L'effort de recensement, qualification, validation des cas d'usage doit également intégrer le niveau de maturité / confiance des analyses envisagées : Par exemple, certaines études préalables exploratoires peuvent conclure à l'absence de modèles d'analyse intéressants / pertinents et être abandonnées à l'issue des premiers travaux d'exploration.

Fabrice DESCHAMPS, COVEA

3.4 PILOTER L'IMPLÉMENTATION DE LA STRATÉGIE DATA ET SUIVRE SES BÉNÉFICES

Les entreprises et administrations publiques cherchent à suivre la mise en œuvre de la stratégie data d'une part et à identifier et mesurer ses bénéfices dans les métiers et l'écosystème d'autre part. Pour cela, les participants au groupe de travail « Gouvernance et architecture data & analytics » ont partagé les indicateurs utilisés.

¹⁶ Proof of Concept : preuve du concept, de la faisabilité.

¹⁷ Minimum Viable Product

3.4.1 PILOTER LA MISE EN ŒUVRE DE LA STRATÉGIE DATA.

L'avancement de la mise en œuvre de la stratégie se mesure sur différents axes : l'organisation, la gouvernance, les processus, les outils et la capacité.

3.4.1.1 Organisation

Plusieurs indicateurs ou mesures ont été partagés par les entreprises participantes.

- L'indicateur de **satisfaction des utilisateurs ou clients**. Il permet d'évaluer la satisfaction et la fidélité d'un client à un moment et de suivre l'évolution du rapport client/marque. Il peut s'agir de clients internes, des métiers, de clients externes au niveau national ou international ou des employés. En fonction du résultat, l'entreprise choisit ou non de mettre en place un plan d'action.
- Le **nombre de produits en cours de développement ou en service**. Cela permet de suivre sur l'aspect technique, la capacité des infrastructures dans la mise en place de la stratégie data.
- Un **abaque modélisant les ressources avec tous les rôles** autour de la gouvernance data et basé sur les processus de l'entreprise. Cela permet de visualiser l'adéquation des nominations et des ETP (équivalent temps plein). Une fois l'abaque calibré, l'entreprise identifie les équipes à renforcer.
- Un indicateur sur la **taille du centre de service data**. Une entreprise préconise de créer un service qui s'occupe de la data pour suivre tous les projet mis en œuvre dans lesquels les données sont utilisées. Cet indicateur permet de voir l'évolution du nombre de projets impliquant la data et de recruter les collaborateurs en conséquence ;
- Le **nombre de responsables identifiés de la donnée**. In fine, l'entreprise doit pouvoir identifier tous les responsables des données. Dans une des entreprises participantes, une fois le responsable de la data identifié, une lettre de mission lui est envoyée officiellement.

3.4.1.2 Gouvernance

Plusieurs indicateurs ou mesures sur la gouvernance ont été partagés par les entreprises participantes.

- Le **nombre d'utilisateurs du catalogue data**. Il s'agit pour cet indicateur d'une métrique mesurant le nombre de consultations du catalogue data, avec le KPI (*Key Performance Indicator*) du déploiement,
- La **proportion du SI documentée dans le catalogue des données**, autrement dit le taux de couverture du catalogue data ;
- Le **volume de données stockées, et ayant été traitées** ;
- Le **nombre de requêtes de données gérées par mois**, avec un ratio entre les différents domaines.
- La **proportion (pourcentage) de projets d'évolution du SI** mettant en œuvre les bonnes pratiques de la gouvernance data.

- Des entreprises ont mis en place une **note de la maturité de la donnée**. Cela se fait par exemple avec une demande de feedback afin que les utilisateurs des données donnent leur avis sur leur qualité et disponibilité ;
- Le **nombre d'indicateurs utilisés pour qualifier la qualité des données**. Les entreprises commencent par ceux qui sont critiques pour le marché, par exemple les données du CRM (*Customer Relationship Management*, i.e. Gestion de la relation clients) utilisées pour le *business* et le marketing. C'est bien que les champs du mail et du portable soient remplis. C'est mieux si le mail et le numéro sont les bons. Ainsi, plusieurs indicateurs¹⁸ ont été partagés :
 - La **complétude** (*Completeness*) ;
 - La **validité** (*Validity*) ;
 - La **précision**, (*Accuracy*) ;
 - La **cohérence** (*Consistency*) ;
 - La **fraîcheur** qui mesure le temps entre le moment où les données sont attendues et celui où elles sont disponibles (*Timeliness*) ;
 - L'**actualité** Plus les données sont fraîches, et plus elles reflètent l'état réel –à un instant T- du concept. (*Currency*) ;
 - L'**accessibilité** (*Accessibility*) ;
 - L'**intégrité**, (*Integrity*) ;
 - La **précision** (*Precision*). Cet indicateur évalue la précision et donne la granularité de la donnée ;
- Le **lignage** (*Lineage*) apporte des informations sur l'origine de la donnée et sa transformation ;
 - La **représentation** de la donnée ;
 - L'**unicité**.

Une entreprise utilise le Lean 6 Sigma¹⁹ dans ses cas d'usage data. Ainsi, une fois le type de dysfonctionnement identifié dans un cas d'usage, l'arbre de décision permet de résoudre le problème avec la mesure de différents indicateurs pour suivre la correction du dysfonctionnement.

Plusieurs participants préconisent de centraliser les investissements liés à la data afin d'avoir une vision globale et cohérente au niveau du groupe.

3.4.1.3 Processus / outils / capacité du SI

Les participants ont partagé plusieurs indicateurs ou mesures sur les processus et les outils utilisés dans leur organisation :

- Le **niveau de déploiement du CI/CD de la data**. L'approche CI/CD (intégration continue/distribution continue et/ou déploiement continu) permet d'augmenter la fréquence de distribution des applications grâce à l'introduction de l'automatisation au niveau des étapes de développement des applications. L'approche CI/CD garantit une automatisation et une surveillance continue tout au long du cycle de vie des applications, des phases d'intégration et de test, jusqu'à la distribution et au déploiement ;

¹⁸ Les indicateurs sont décrits sur ce site : <http://dimensionsofdataquality.com/content/list-underlying-concepts>

¹⁹ <https://leansixsigmafrance.com/blog/la-difference-entre-le-lean-le-six-sigma-et-le-lean-six-sigma/>

- **La disponibilité d'outils pour mettre à disposition** la data et l'industrialisation de ces outils ;
- Le **nombre de produits** ou outils **data disponibles** ;
- **Le nombre de cas d'usage** ;
- La durée du **processus de qualification d'un cas d'usage data** de l'idéation initiale, phase d'exploration permettant de confirmer ou non la pertinence d'un cas d'usage, jusqu'à la mise en production effective. La mise en place d'un processus de qualification des nouveaux cas d'usage constitue une des avancées de la démarche itérative de déploiement de la stratégie data ;
- Le **taux d'automatisation ou d'industrialisation de la donnée de bout en bout de sa chaîne de traitement** (ou alors les niveaux d'interopérabilité des "modules" Data) - Les outils de la chaîne de traitement de la donnée incluent : le catalogue des données, les référentiels de gestion des données (Master Data Management), des ETL (*Extract, Transform and Load*²⁰), des API, de la BI (*Business Intelligence*) et de l'IA (*Intelligence artificielle*).

Faut-il choisir un outil complexe avec une multitude de modules compliqués à mettre en œuvre et à interopérer avec les autres outils de la chaîne data, ou bien, faut-il plutôt choisir un outil avec moins de fonctionnalités mais simple à interconnecter dans la chaîne ? Le *business* donne sa préférence aux outils qui promettent le plus de fonctionnalités. Cependant, les participants s'accordent sur l'importance de privilégier les outils assurant le traitement de la donnée dans sa largeur, c'est-à-dire de bout en bout afin d'en retirer le maximum de valeur (couverture verticale). À cela il faut également ajouter l'indicateur mesurant la capacité d'un outil à traiter des cas d'usage en parallèle (couverture horizontale). Les participants notent que les premiers cas d'usage posent beaucoup de difficultés sur l'intégration verticale et horizontale mais une fois la méthodologie mise en place, les suivants sont plus simples. Une entreprise utilise l'outil Lean 6 Sigma pour cela.

3.4.2 MESURER LES BÉNÉFICES BUSINESS ET LA VALEUR APPORTÉS PAR L'USAGE DE LA DATA

Une bonne pratique est de labelliser les indicateurs afin de les réutiliser dans plusieurs entités de l'entreprise de manière équivalente. Cependant, il faut être vigilant sur ce type de données. En effet, l'indicateur est une donnée qui, vue de loin, peut sembler commune à de nombreux métiers dans l'entreprise, mais qui se révèle, en y regardant de plus près, porter des spécificités en fonction des métiers. La partager nécessite donc un travail préalable d'alignement, pas toujours facile. C'est pourquoi, l'objectif est de labelliser les indicateurs pour qu'ils puissent être réutilisés au sein de l'entreprise, mais avec la même base de calcul. **L'intérêt est que tout le monde parle le même langage.**

Afin d'acculturer les collaborateurs du *business* et de l'exécutif à la stratégie data et de les embarquer, il est important d'en démontrer la valeur par des premiers résultats concrets et convaincants.

Le groupe de travail « Gouvernance et architecture data & analytics » a partagé les indicateurs utilisés pour mesurer les bénéfices apportés par la stratégie data elle-même dans les métiers et l'écosystème. Ces bénéfices s'articulent autour du *time to market*, de la performance et du chiffre d'affaires.

²⁰ Ce processus d'intégration revient à extraire les données brutes d'un système, à les transformer et à les envoyer vers un entrepôt de référence. Il vise à réaliser des synchronisations massives entre différents systèmes. Les outils ETL sont d'excellents moyens pour collecter des données de différentes provenances puis les exploiter. Ils représentent des solutions phares du Big Data et de la *Business Intelligence*.

3.4.2.1 Time to market

- Le **nombre de PoC**. Une entreprise s'est par exemple fixé un objectif en nombre de PoC de cas d'usage data ;
- Le **coût et la durée des PoC** ;
- La **vitesse de passage, de l'expérimentation du PoC jusqu'à la production** : une fois le PoC validé, une entreprise s'est donnée pour objectif de passer à la mise en production en 6 mois. Elle a dû pour cela simplifier le processus alors trop lourd.
- Le **gain de temps** obtenu grâce à l'implémentation d'un cas d'usage data.

3.4.2.2 Performance

- La **réduction du nombre d'ETP** (Équivalent Temps Plein) observée suite à la mise en place du cas d'usage data qui a permis d'éviter des interventions inutiles, ou de réduire le temps de traitement, par exemple ;
- La **réduction des risques identifiés** ;
- Le **nombre d'API** qui assure la récupération rapide des données avec une analyse du marché. Une entreprise achète à un fournisseur externe la possibilité d'importer des données fiables en open data afin de compléter son CRM.

3.4.2.3 Chiffre d'affaires

- Le **nombre de nouveaux services créés** ;
- Le **KPI du retour sur investissement** (ROI : *Return Of Investment*). Une réduction conséquente du ROI suite à la mise en place d'un cas d'usage data contribue à embarquer le *business* dans la stratégie data de l'entreprise ;
- La **conformité réglementaire** (par exemples sur le RGPD ou sur la qualité). Le risque de non-conformité expose à des conséquences financières. Pour cela il est important de faire prendre conscience du triptyque « Risque-Incident-Conséquence » de la non-conformité.

4 CONCLUSION

La plupart des entreprises et administrations publiques génèrent une quantité très importante de données, que ce soit des données régaliennes (finances, RH, réglementaires...), ou des données liées à leur activité. Une partie plus ou moins importante de l'exploitation de ces données est aujourd'hui réalisée de manière atomisée à travers des initiatives hors de la DSI. Certaines organisations cherchent à ouvrir leur SI avec une approche de services adressés à l'ensemble de son écosystème. Cette capacité à partager les données à l'échelle pour générer de la valeur, constitue un critère de compétitivité.

On constate une conviction commune aux entreprises et administrations publiques quant au caractère stratégique des données. Néanmoins, cette conviction reste diversement partagée au sein de chacune des organisations, matérialisant des maturités collectives encore hétérogènes. L'appréciation de la maturité collective autour de la « culture de la donnée » au sein des organisations est essentielle pour établir une trajectoire raisonnable et réaliste visant à décliner une stratégie autour de la donnée. La stratégie data de l'entreprise ne constitue pas un sujet technologique mais un sujet de **transformation d'entreprise** dans laquelle **l'ensemble des équipes a un rôle à jouer**. En fonction de la stratégie data « retenue », le sujet technologique se traitera différemment. L'acculturation des équipes aux données est primordiale et la DSI est appelée à exercer une responsabilité majeure qui s'articule autour de plusieurs actions :

- **Définir et structurer la démarche de déploiement d'un cadre de gouvernance** des données de sa propre initiative, si les directions métier ne sont pas sensibles ni motrices en la matière ;
- **Anticiper la nécessaire industrialisation des bonnes pratiques**, par la dotation de capacités outillées (ex : catalogue des données) ;
- **Faciliter l'appropriation du patrimoine des données et l'émergence de cas d'usage**, par la confection d'offres de services favorisant l'expérimentation en « mode lab ».

Cette exploitation accrue des données au service du métier constitue une lame de fond inéluctable, appelée à s'amplifier avec le foisonnement de nouvelles natures de données et la généralisation des échanges de données inter-organismes (ex : [loi 3DS](#)). En réponse aux limites montrées par une trop grande centralisation dans les *data lakes*, et aussi pour positionner des traitements performants au plus près de la source, certaines entreprises ou administrations publiques mettent en place une logique de réseau de données (*data mesh*).

Par ailleurs, la stratégie data, notamment dans sa gouvernance, doit prendre en compte l'impact écologique qui devient de plus en plus prégnant. En effet, les données représentent à la fois une problématique et un des éléments de réponse au défi de sobriété. Tout d'abord, l'augmentation exponentielle du volume de données issues de capteurs et de systèmes complexes qu'il a fallu produire et qu'il faudra recycler, mais également, leur stockage, leur partage et leur traitement analytique ont des répercussions sur l'empreinte carbone de l'entreprise ou de l'administration publique²¹. Cependant, les données elles-mêmes sont de précieux outils d'information et de pédagogie pour réduire l'empreinte carbone de l'ensemble des activités de l'organisation et aider à adopter une démarche environnementale responsable et à l'améliorer étape par étape, en fonction des objectifs définis.

²¹ Cf. le rapport Cigref, en partenariat avec The Shift Project, « Sobriété numérique : une démarche d'entreprise responsable » librement téléchargeable : https://www.cigref.fr/publication-sobriete-numerique-une-demarche-d-entreprise-responsablequi_propose_100_bonnes_pratiques_de_sobriete_numerique.



Au service de la croissance économique et de la compétitivité de nos membres, grandes entreprises et administrations publiques françaises, utilisatrices de solutions et services numériques, par la réussite du numérique

Le Cigref est un réseau de grandes entreprises et administrations publiques françaises qui a pour mission de développer la capacité de ses membres à intégrer et maîtriser le numérique. Par la qualité de sa réflexion et la représentativité de ses membres, il est un acteur fédérateur de la société numérique. Association loi 1901 créée en 1970, le Cigref n'exerce aucune activité lucrative.

Pour réussir sa mission, le Cigref s'appuie sur trois métiers, qui font sa singularité.

Appartenance

Le Cigref incarne une parole collective des grandes entreprises et administrations françaises autour du numérique. Ses membres partagent leurs expériences de l'utilisation des technologies au sein de groupes de travail afin de faire émerger les meilleures pratiques.

Intelligence

Le Cigref participe aux réflexions collectives sur les enjeux économiques et sociétaux des technologies de l'information. Fondé il y a près de 50 ans, étant l'une des plus anciennes associations numériques en France, il tire sa légitimité à la fois de son histoire et de sa maîtrise des sujets techniques, socle de compétences de savoir-faire, fondements du numérique.

Influence

Le Cigref fait connaître et respecter les intérêts légitimes de ses entreprises membres. Instance indépendante d'échange et de production entre praticiens et acteurs, Il est une référence reconnue par tout son écosystème.

www.cigref.fr
21 av. de Messine, 75008 Paris
+33 1 56 59 70 00
cigref@cigref.fr