Haute Ecole

« ICHEC – ECAM – ISFSC »

Enseignement supérieur de type long de niveau universitaire

Amélioration des processus de gestion des données au sein de Peppl. et proposition d'un programme de gouvernance des données adapté à l'entreprise

Mémoire présenté par :

Sylvain Gouverneur

Pour l'obtention du diplôme de :

Master - Ingénieur commercial

Année académique 2023-2024

Promoteur :

Alain Ejzyn

Boulevard Brand Whitlock 6 - 1150 Bruxelles

Remerciements

Tout d'abord, je remercie chaleureusement mon promoteur, Alain Ejzyn, pour sa disponibilité et ses conseils judicieux ayant permis d'orienter ce projet et d'améliorer mon approche de recherche.

Je tiens également à exprimer ma reconnaissance envers l'équipe de Peppl. pour leur accueil chaleureux et pour m'avoir fait sentir intégré au sein de l'entreprise. Je remercie tout particulièrement ma maître de stage, June Van Veer, pour sa confiance et son soutien tout au long de mon stage.

Je souhaite aussi remercier mes parents, Isabelle Leroy et Bruno Gouverneur, pour leur aide précieuse tout au long du projet et pour la relecture attentive de ce mémoire.

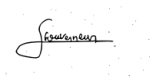
Je tiens également à exprimer ma gratitude à Benjamin Abdi pour l'entretien qu'il m'a accordé, lequel a contribué à l'enrichissement de certains aspects de ce mémoire.

Enfin, je tiens à remercier encore une fois toutes les personnes qui ont, de près ou de loin, contribué à la réalisation de ce mémoire.

Engagement Anti-Plagiat du Mémoire

« Je soussigné, GOUVERNEUR Sylvain, Master ingéCo année terminale, déclare par la présente que le travail ci-joint respecte les règles de référencement des sources reprises dans le règlement des études en signé lors de mon inscription à l'ICHEC (respect de la norme APA concernant le référencement dans le texte, la bibliographie, etc.) ; que ce travail est l'aboutissement d'une démarche entièrement personnelle ; qu'il ne contient pas de contenus produits par une intelligence artificielle sans y faire explicitement référence. Par ma signature, je certifie sur l'honneur avoir pris connaissance des documents précités et que le travail présenté est original et exempt de tout emprunt à un tiers non-cité correctement. »

Date : 18 août 2024 Signature :



Je soussigné(e), GOUVERNEUR Sylvain 190813 (nom + numéro de matricule), déclare sur l'honneur les éléments suivants concernant l'utilisation des intelligences artificielles (IA) dans mon travail / mémoire :

Type d'assistance		**Case à cocher**
Aucune assistance	J'ai rédigé l'intégralité de mon travail sans avoir eu recours à un outil d'IA générative.	
Assistance avant la rédaction	J'ai utilisé l'IA comme un outil (ou moteur) de recherche afin d'explorer une thématique et de repérer des sources et contenus pertinents.	X
Assistance à l'élaboration d'un texte	J'ai créé un contenu que j'ai ensuite soumis à une IA, qui m'a aidé à formuler et à développer mon texte en me fournissant des suggestions.	
	J'ai généré du contenu à l'aide d'une IA, que j'ai ensuite retravaillé et intégré à mon travail.	X
	Certains parties ou passages de mon travail/mémoire ont été entièrement été générés par une IA, sans contribution originale de ma part.	
Assistance pour la révision du texte	J'ai utilisé un outil d'IA générative pour corriger l'orthographe, la grammaire et la syntaxe de mon texte.	X
	J'ai utilisé l'IA pour reformuler ou réécrire des parties de mon texte.	X
Assistance à la traduction	J'ai utilisé l'IA à des fins de traduction pour un texte que je n'ai pas inclus dans mon travail.	X
	J'ai également sollicité l'IA pour traduire un texte que j'ai intégré dans mon mémoire.	
Assistance à la réalisation de visuels	J'ai utilisé une IA afin d'élaborer des visuel, graphiques ou images.	
Autres usages		

Je m'engage à respecter ces déclarations et à fournir toute information supplémentaire requise concernant l'utilisation des IA dans mon travail / mémoire, à savoir :

J'ai mis en annexe les questions posées à l'IA et je suis en mesure de restituer les questions posées et les réponses obtenues de l'IA. Je peux également expliquer quel le type d'assistance j'ai utilisé et dans quel but.

Fait à Bruxelles (ville), le 18 août 2024 (date)

Signature : Sylvain Gouverneur 190813 [Prénom Nom de l'étudiant(e) et matricule]

Table des matières

Liste des Figures

Liste des tableaux

Introduction

"Les données sont le nouveau pétrole." C'est par cette désormais célèbre citation que l'entrepreneur anglais Clive Humby, dès 2006, soulignait l'importance des données pour les entreprises (Haupt, 2016). En 2024, cette comparaison est plus pertinente que jamais, alors que les entreprises déploient des moyens considérables pour collecter des données et en tirent des bénéfices significatifs grâce à une exploitation optimale.

Les données brutes, tout comme le pétrole, ont peu de valeur intrinsèque. Cependant, une fois raffinées et analysées, elles permettent aux entreprises de générer des informations capitales, contribuant à l'amélioration continue de leurs opérations et de leur prise de décision.

Cette importance des données en entreprise s'explique notamment par l'expansion continue du trafic de données mobiles dans le monde. Le cabinet de conseil stratégique Arthur D. Little estime cette croissance annuelle à environ 25% par an en Europe (Burkhanov et al., 2023). Cette progression constante du volume du trafic des données couplée avec l'amélioration des technologies de suivi, d'analyse et de modélisation des données ont propulsé la nécessité d'une exploitation optimale des données au centre des préoccupations des entreprises.

Dans ce contexte où les données occupent une place de plus en plus centrale, de nombreuses entreprises jugent indispensable de mettre en place un ensemble de pratiques, de mesures et de politiques visant à garantir la qualité, la sécurité, et l'utilisation optimale des données au sein de leur organisation.

Cet enjeu, la startup Peppl., qui propose une application axée sur le bien-être mental en est bien consciente. En améliorant les différents processus de gestion des données dont elle dispose, elle s'assurerait une utilisation plus optimale de ses données.

Au-delà de l'examen de la mise en place de ces différents processus, ce mémoire explore de nouvelles pistes pour intégrer ces processus dans un cadre de gouvernance des données. La volonté est de consolider ainsi les bases d'une gestion structurée et centralisée de l'ensemble des données de l'entreprise.

La question centrale à laquelle ce mémoire cherche à répondre est :

"Comment améliorer les processus de gestion de données au sein de la startup Peppl. en les intégrant à un cadre de gouvernance des données?"

La première partie de ce mémoire propose une revue de la littérature qui met en lumière l'évolution de l'importance et de l'utilisation des données digitales en entreprise au fil des années. Elle clarifie le concept d'actif de données et examine en profondeur les différentes étapes du cycle de vie des données. En outre, cette section développe le concept de gouvernance des données et propose un

modèle détaillé des éléments essentiels à mettre en place pour établir un cadre de gouvernance des données efficace au sein d'une entreprise.

Une fois le cadre théorique établi, le mémoire se penche sur le contexte de l'entreprise dans lequel le projet de recherche a été mené. Ce chapitre aborde également les objectifs visés par la gestion de projet ainsi que la méthodologie appliquée pour les atteindre.

a supprimé: ra

La troisième partie du mémoire décrit l'application concrète de cette méthodologie. Cette description fournit une explication détaillée des différents processus mis en place.

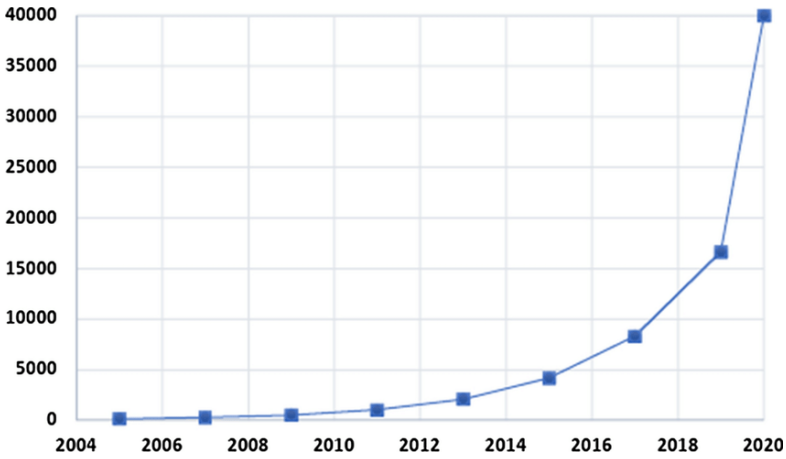
Enfin, la dernière partie du mémoire propose un modèle de gouvernance des données adapté à Peppl.. La proposition fait la synthèse entre les recherches menées pour intégrer les processus développés au cours du projet dans le but d'optimiser l'utilisation des données disponibles au sein de l'entreprise Peppl.

Partie 1 : Contexte théorique du mémoire

1.1 Évolution des données digitales à travers les années

Depuis plusieurs années, le digital occupe une place centrale dans notre société, rendant l'utilisation des technologies digitales indispensable à son bon fonctionnement. Comme l'illustre la Figure 1, l'intégration rapide du virtuel dans nos vies s'accompagne d'une croissance exponentielle du volume de données digitales générées à l'échelle mondiale. Alors qu'au début des années 2000, seules quelques exabytes de données étaient produites chaque année, ce chiffre s'est multiplié pour atteindre plusieurs dizaines de milliers d'exabytes au début des années 2020 (Anter et al., 2021).

Figure 1: Croissance annuelle du volume de données digitales en exabytes dans le monde entre 2005 et 2020



Source : Journal of Big Data (2021). *A graph-based big data optimization approach using hidden Markov model and constraint satisfaction problem*.
https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00485-z

Avant d'essayer de comprendre comment ces données digitales sont devenues un enjeu majeur pour les entreprises, il est utile de retracer leur évolution au fil des décennies. Dans les années 1970, les données digitales étaient principalement vues comme des enregistrements nécessaires pour les opérations internes et la comptabilité (Watson, 2019).

En 1974, les travaux de R. Peterson ont marqué un tournant décisif en reconnaissant l'importance des données économiques pour l'analyse et la prise de décision, notamment à travers l'utilisation des

données transactionnelles pour examiner la demande monétaire aux États-Unis dans son article "A Cross Section Study of the Demand for Money: The United States, 1960-62" (Press, 2019).

Avec l'émergence des technologies de Business Intelligence (BI) et des entrepôts de données dans les années 1990, les entreprises ont commencé à comprendre le potentiel stratégique que représentent les données.

Les technologies de Business Intelligence englobent un ensemble de technologies de soutien à la décision, permettant aux entreprises de convertir des données brutes en informations exploitables pour prendre des décisions éclairées (Chaudhuri & al., 2011).

Un entrepôt de données est un système de stockage centralisé qui intègre, organise et conserve de grandes quantités de données provenant de diverses sources au sein d'une entreprise. Cet entrepôt constitue un élément fondamental de la Business Intelligence, car il fournit la base de données structurée sur laquelle reposent les outils et technologies de BI, facilitant ainsi l'analyse et la prise de décisions stratégiques (Aljuwaiber, 2022).

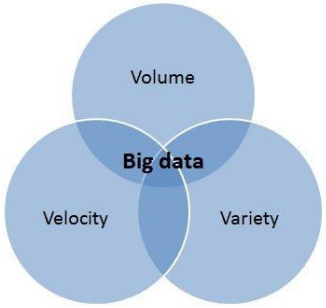
Des ouvrages influents, tel que "Working Knowledge: How Organizations Manage What They Know" de T. Davenport et L. Prusak (1998), ont souligné l'importance de transformer les données en informations pertinentes et en connaissances, redéfinissant les données comme des "faits objectifs et discrets sur des événements" (Press, 2019).

En parallèle, comme souligné par R.J.T Morris et B.J. Truskowski (1996), la réduction des coûts de stockage digital par rapport au papier a permis une meilleure gestion des volumes croissants de données. Des termes comme "big data" ont émergé pour décrire les défis liés à la gestion de grandes quantités de données, popularisés par des publications et des présentations de chercheurs comme Michael Cox, David Ellsworth et John R. Mashey à la fin des années 1990 (Press, 2019).

En 2011, l'IDC (International Data Corporation) définit le big data comme "une nouvelle génération de technologies et d'architectures, conçues pour extraire économiquement la valeur de très grands volumes de données variées, permettant une grande vitesse de capture, de découverte et/ou d'analyse" (cité par Lei & Kong, 2020, p.30), et ce concept est rapidement devenu central dans le domaine de l'informatique et des affaires.

En 2001, l'analyste Doug Laney a introduit la formulation des "3 Vs" pour décrire les dimensions clés du big data dans son article " “3D Data Management: Controlling Data Volume, Velocity, and Variety." (Press, 2019).

Figure 2 : Les 3 dimensions du Big data



Source : Optalitix (2020, Janvier 23) What are the 3 V's of big data? https://www.optalitix.com/insights/what-are-the-3-vs-of-big-data

Comme indiqué sur la Figure 2, les 3 V se réfèrent au :

- **Volume** : La quantité massive de données générées.
- **Vélocité** : La vitesse à laquelle ces données sont produites et traitées.
- **Variété** : La diversité des types de données disponibles.

Ce cadre a contribué à une meilleure compréhension des défis et opportunités liés aux big data.

En 2009, l'Association Internationale de Management des Données a reconnu les données comme des actifs essentiels pour les entreprises. Les recherches, telles que celles de T. Fisher dans son ouvrage "The Data Asset: How Smart Companies Govern Their Data for Business Success" (2009), ont ensuite mis en lumière la valeur stratégique durable des données, les qualifiant de ressources à long terme et non épuisables pour les organisations (Shi et al., 2023).

En effet, à la différence des ressources physiques, les actifs de données ne se dégradent ni avec le temps ni avec l'usage. Elles peuvent être exploitées et réutilisées indéfiniment, tant qu'elles sont correctement conservées, et ne s'épuisent pas, quel que soit le nombre de fois où elles sont utilisées.

Depuis 2010, l'essor rapide des technologies de big data, de l'Internet des objets (IoT), du cloud computing et de l'intelligence artificielle a considérablement augmenté les capacités de traitement des données. Cette évolution a permis aux entreprises de percevoir les données non seulement comme une ressource précieuse, mais aussi comme un actif stratégique offrant un avantage compétitif significatif (Hou & Xiao, 2019).

En parallèle, sous l'impulsion des travaux de chercheurs et experts comme Lukoianove T. et Rubin V. avec "Veracity Roadmap: Is Big Data Objective, Truthful and Credible?" (2014), la composante de la Véracité a été ajoutée aux trois dimensions initiales du Big data (Volume, Vélocité, Variété). Cette quatrième dimension souligne l'importance de la qualité et de la fiabilité des données pour garantir

des analyses précises et des décisions éclairées, répondant ainsi aux défis croissants liés à la gestion de données massives et diverses (Press, 2019).

Ainsi, l'évolution intervenue dans la perception des données, passant d'un simple outil de gestion interne à un actif stratégique crucial, illustre l'importance croissante de l'analyse et de la gestion des données pour les entreprises.

1.2 Actifs de données en entreprise

Comme nous l'avons vu dans la section précédente, la valeur des données n'a cessé de croître au fil des années, au point de devenir une ressource essentielle pour les entreprises, comparable en importance aux ressources humaines et financières. Selon Rajnoha et Hada (2021), il est impératif pour les entreprises de capitaliser sur ces actifs de données afin de créer de la valeur et d'obtenir un avantage compétitif par rapport à leurs concurrents. Bruman (2023) souligne également que, dans le monde numérique actuel, la capacité à prendre des décisions fondées sur les données collectées et à élaborer des stratégies basées sur leur analyse est primordiale pour assurer le succès des entreprises.

1.2.1 Définition d'un actif de données

Tout au long de ce mémoire, on qualifie en tant qu'actif les données dont dispose l'entreprise. Il est donc essentiel de définir ce que recouvre précisément ce terme d'actif de données.

Il n'existe pas de consensus sur la définition d'un actif de données, mais on peut par exemple se tourner vers la définition proposée par l'institut des informations et communications de Chine qui définit les actifs de données comme "des ressources de données enregistrées physiquement ou électroniquement, telles que des documents et des données électroniques, qui sont détenues ou contrôlées par une entreprise et peuvent apporter des bénéfices économiques futurs à l'entreprise" (cité par CAICT, 2019, p. 28).

On peut également se pencher sur la définition élaborée dans l'article "From data to data asset: conceptual evolution and strategic imperatives in the digital economy era" qui définit les actifs de données comme des "ressources de données détenues ou contrôlées par une entreprise et qui ont une valeur réelle ou potentielle, sont conformes aux lois sur les données et sont enregistrées électroniquement" (Shi et al., 2023, p. 6).

Ces définitions très similaires abordent le concept d'actif de données sous plusieurs aspects :

Possession et contrôle

En premier lieu, on a l'actif de données en tant que ressources de donnée possédée ou contrôlée par une entreprise, ce qui signifie qu'elle en dispose de telle manière à atteindre ses objectifs stratégiques,

pour autant qu'elle respecte les lois sur les données qui veillent à la protection et l'intégrité des données récoltées, comme l'indique la deuxième définition, plus récente.

Apport

D'autre part, les données apportent un bénéfice potentiel et réel à l'entreprise, c'est-à-dire qu'elles vont soit directement bénéficier à l'entreprise, au niveau des opérations actuelles, soit elles pourront potentiellement lui apporter un avantage dans le futur.

Electronique/physique

Enfin, une définition propose de regrouper sous le terme d'actifs de données les données enregistrées électroniquement ou physiquement, tandis que l'autre considère uniquement les données digitales. La première approche, soutenue par H. Hannila et al. (2019), reconnaît que les enregistrements papier peuvent encore contenir des informations significatives et avoir de la valeur dans des contextes spécifiques. La deuxième approche, devenue la norme depuis, soutient la commodité et la facilité de gestion en lien avec la forme des données car les données enregistrées électroniquement sont généralement plus fiables et accessibles que celles qui ne le sont pas. Ainsi, plusieurs universitaires estiment que la valeur des données doit être réalisée sur la base d'analyses à grande échelle, nécessitant ainsi des enregistrements électroniques (Shi et al., 2023).

1.2.2 Apport des actifs de données

Les actifs de données peuvent avoir un impact positif sur différents aspects de l'entreprise, influençant ainsi plusieurs domaines de son fonctionnement. Ils vont par exemple fournir des informations vitales dans le cadre de la prise de décisions opérationnelles à travers des indicateurs de performance clés (KPIs) et d'autres données pertinentes qui évaluent les performances de l'entreprise en temps réel (Shi et al., 2023).

Dans le cadre de la planification stratégique, les actifs de données offrent des informations sur les tendances du marché et les forces concurrentielles. Ces informations aident ainsi à formuler des stratégies alignées sur les dynamiques du marché, à allouer efficacement les ressources disponibles et à mettre en place des objectifs réalistes (Shi et al., 2023).

Les actifs de données sont également très importants pour le développement de nouveaux produits et pour l'amélioration de ceux déjà existants. Ils permettent tout d'abord de saisir la performance des produits existants, en offrant des perspectives sur les expériences des utilisateurs et sur la performance des produits. Cette compréhension approfondie alimente une amélioration continue de ces derniers, renforçant ainsi la satisfaction des utilisateurs et la valeur sur le marché. Enfin, les données guident le processus de développement des nouveaux produits en fournissant des informations provenant du marché et des attentes des consommateurs (Shi et al., 2023).

1.2.3 Distinction des différents actifs de données

Parmi les divers actifs de données existants, il est important de saisir les distinctions entre les différents types de données existants afin de comprendre pleinement les spécificités propres à chaque type.

Il y a différentes façons de catégoriser les actifs de données. On peut par exemple effectuer des distinctions basées sur la provenance des données, sur leur nature ou sur leur utilisation dans les processus opérationnels.

Catégorisation basée sur la nature des données

Les données se répartissent en deux catégories selon leur nature, qu'elle soit qualitative ou quantitative.

Les **données quantitatives** sont des données numériques pour lesquelles les relations d'ordre et de proportionnalité ont un sens, comme le salaire, l'âge ou le nombre d'heures travaillées dans la semaine.

Les **données qualitatives**, elles, peuvent être soit nominales (renseignées par des lettres), soit numériques mais sans relations d'ordre et de proportionnalité significatives. Comme exemple de données qualitatives on peut citer le genre ou le statut civil. Ces variables ne permettent pas de déterminer un ordre ou une proportion significative entre les différentes catégories (Coron, 2020).

Catégorisation basée sur la provenance des données

Une autre manière de différencier les données provenant de consommateurs est de les séparer en données de première partie et en données provenant de tiers.

Les **données de première partie** sont des données que l'organisation recueille directement auprès des utilisateurs lorsque des derniers recourent aux technologies faisant partie de son infrastructure digitale (site web, application, réseaux sociaux,...). Ces informations sont les plus utiles pour les entreprises pour plusieurs raison (Cote 2021).

Selon J. Horn et B. Bruno (2022), les données collectées de cette manière sont exhaustives, car elles capturent l'ensemble des comportements et interactions des utilisateurs sur les canaux. Elles sont précises, car contrairement aux données provenant de tiers, elles ne sont pas agrégées à partir de différentes sources.

De plus, elles sont pertinentes, car elles proviennent des canaux propres à l'entreprise, et fournissent ainsi des informations précieuses sur le comportement des utilisateurs par rapport aux produits de l'entreprise. Il y a également un avantage économique significatif à utiliser ces données, car elles ne nécessitent pas de paiement nécessaire à leur obtention, contrairement aux données provenant de tiers.

Enfin, en les collectant de manière autonome, l'entreprise dispose d'un contrôle total sur ces données, ce qui renforce la confidentialité et la sécurité, tout en offrant une meilleure opportunité d'analyse et de personnalisation (Horn & Bruno, 2022).

Les **données de tierces parties**, quant à elles, sont collectées par des entités extérieures à l'entreprise. Elles proviennent généralement de multiples sources, sont agrégées et souvent anonymisées. L'accès à ces données est généralement payant (Cote 2021).

L'agrégation des données est le processus de compilation et de combinaison de données individuelles provenant de différentes sources pour les organiser de manière à faciliter leur analyse et leur interprétation. L'agrégation des données permet de masquer les détails individuels pour se concentrer sur les tendances générales, les moyennes, ou d'autres mesures statistiques globales (Egidius et al., 2019).

Les plateformes publicitaires telles que Facebook ou Google sont des exemples courants de fournisseurs de données de tierces parties, offrant des informations sur les types d'audience ayant interagi avec les publicités de l'entreprise sur leurs canaux.

Bien que ces données soient moins précises et spécifiques que les données de première partie, elles permettent aux entreprises utilisant ces plateformes d'obtenir une vue plus large et diversifiée du comportement des consommateurs, en particulier lorsqu'elles sont agrégées de multiples sources et analysées en profondeur (Cote 2021).

Catégorisation basée sur l'utilisation des données dans les processus opérationnels

Une dernière manière de différencier les actifs de données est celle développée dans l'article "Data-driven begins with DATA; Potential of data assets" (Hannila et al., 2019), dans lequel les données sont séparées en données maîtres, données transactionnelles et données d'interaction.

Les **données maîtres** sont des informations fondamentales et stables qui décrivent les entités essentielles d'une entreprise, telles que les produits, les clients et les fournisseurs. Ces données restent relativement constantes tout au long du cycle de vie de ces entités et servent de référence commune pour divers systèmes et processus au sein de l'organisation. Par exemple, les données sur les produits incluent les descriptions, les catégories, les prix et les spécifications techniques, tandis que les données sur les clients contiennent des informations démographiques, des coordonnées et des historiques d'achat (Hannila et al., 2019).

Les **données transactionnelles** sont générées par les activités commerciales quotidiennes. Elles capturent les interactions et les échanges entre les entités de l'entreprise, comme les commandes, les factures, les paiements et les livraisons. Ces données sont essentielles pour le suivi et la gestion des opérations commerciales, permettant d'analyser les performances financières, de suivre les ventes et les stocks, et de gérer les relations avec les clients et les fournisseurs. Par exemple, les enregistrements

de toutes les transactions de vente effectuées par une entreprise au cours d'une période donnée constituent des données transactionnelles (Hannila et al., 2019).

Les **données d'interaction** proviennent des interactions entre les personnes et les machines. Ces données capturent les comportements, les préférences et les interactions des utilisateurs avec les systèmes et les appareils. Elles sont utiles pour comprendre comment les utilisateurs interagissent avec les produits et services de l'entreprise, permettant d'optimiser les expériences utilisateur et d'améliorer les interfaces et les produits. Par exemple, les logs d'utilisation d'une application mobile enregistrant chaque interaction de l'utilisateur, ou les données collectées par des capteurs IoT sur des machines industrielles surveillant les performances en temps réel, sont des données d'interaction (Hannila et al., 2019).

Ces catégorisations des actifs de données sont essentielles pour les entreprises car elles permettent d'évaluer la qualité des données, de déterminer les méthodes d'analyse et de croisement appropriées, et d'identifier les processus de maintenance nécessaires. En comprenant ces distinctions, les entreprises peuvent optimiser la gestion de leurs données, améliorer la précision de leurs analyses et garantir une utilisation efficace et sécurisée des informations.

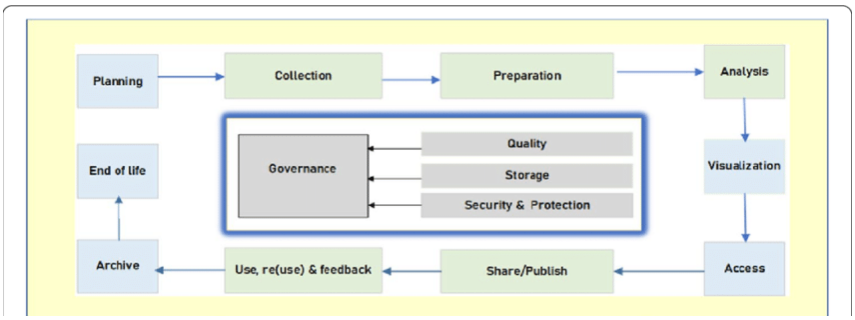
Ces différents actifs de données, quelle que soit leurs caractéristiques, traversent un ensemble d'étapes tout au long de leur parcours au sein de l'entreprise. De leur collecte initiale à leur suppression éventuelle, ces étapes constituent ce que l'on appelle le cycle de vie des données.

1.3 Le cycle de vie des données

On peut définir le cycle de vie des données comme "la séquence d'étapes par lesquelles une unité de données particulière passe, depuis sa génération ou sa capture initiale jusqu'à son archivage éventuel et/ou sa suppression à la fin de sa durée de vie utile" (Kirvan et al., 2023, para. 1).

Il existe de nombreux modèles de cycle de vie des données, chacun comportant un nombre variable de phases et pouvant être adapté à divers contextes, qu'il s'agisse de la recherche ou de la gestion de données en entreprise. Pour ce mémoire, nous utiliserons le modèle DaLiF (Data Lifecycle Framework) (Figure 3), développé par Syed Iftikhar Hussain Shah, Vassilios Peristeras et Ioannis Magnisalis en 2021. Ce modèle est basé sur une synthèse de nombreux modèles préexistants, offrant ainsi une approche compréhensive et flexible pour la gestion du cycle de vie des données (Magnisalis et al., 2021).

Figure 3 : Modèle DaLiF du cycle des données



Source : Journal of Big Data (2021) DaLiF-Data Lifecycle Framework for data-driven governments https://www.researchgate.net/figure/DaLiF-Data-Lifecycle-Framework-for-data-driven-governments_fig3_352382112

Ce modèle comporte 14 phases, dont 10 se déroulent de manière successive et 4 se déroulent tout au long du cycle. Les étapes en vert sont considérées par le modèle comme essentielles au bon fonctionnement du cycle de vie, tandis que celles en bleu sont des phases optionnelles du cycle de vie. Ce modèle offre un aperçu complet des différentes possibilités de gestion du cycle de vie des données. Cette flexibilité permet aux organisations de personnaliser et d'adapter le modèle en fonction de leurs besoins spécifiques et de leurs contextes opérationnels, tout en bénéficiant d'une structure cohérente pour maximiser la valeur et l'efficacité des données à chaque étape de leur cycle de vie (Magnisalis et al., 2021).

1.3.1 Phases successives du cycle de vie des données

1ère phase : La planification

Selon N. Glover (2022), la planification des données implique de prévoir les besoins spécifiques en données, les sources, les méthodes de collecte et de stockage, ainsi que le traitement, la présentation, la distribution et la sécurité des données. Un plan de gestion des données est essentiel pour organiser, gérer, partager et préserver les données de manière efficace.

Les avantages de la planification des données incluent la réduction des risques de vol et de perte de données, l'amélioration de la qualité et de l'intégrité des données, et l'optimisation de l'utilisation des ressources.

En somme, une planification rigoureuse des données permet aux entreprises de maximiser leur valeur tout en assurant leur protection et leur disponibilité à long terme (Glover, 2022).

2ème phase : La collecte

La phase de collecte de données implique l'acquisition de données provenant de diverses sources internes et externes, sous différentes formes (structurées, semi-structurées et non structurées). Cette phase définit le moment ou de la nouvelle donnée entre dans le système (Magnisalis et al., 2021).

Données structurées

Les données structurées sont généralement des données tabulaires, représentées sous forme de colonnes et de lignes dans une base de données. Les bases de données qui stockent ces tables sont appelées bases de données relationnelles. Dans les données structurées, chaque ligne d'une table possède le même ensemble de colonnes (Alam, 2024).

Données semi-structurées

Les données semi-structurées sont des informations qui ne suivent pas le format des données structurées (bases de données relationnelles) mais qui possèdent néanmoins une certaine structure (Alam, 2024).

Données non structurées

Les données non structurées sont des informations qui ne sont ni organisées de manière prédéfinie ni structurées selon un modèle de données établi. Elles sont souvent riches en texte, mais peuvent également contenir des éléments tels que des chiffres, des dates, et des faits. Les vidéos, les fichiers audios, et les données binaires qui n'ont pas de structure spécifique sont classés comme des données non structurées (Alam, 2024).

Les données peuvent provenir de capteurs IoT, de réseaux sociaux, de systèmes d'entreprise, de formulaires en ligne, d'enquêtes ou de sources publiques.

Pour collecter les données provenant de différentes sources, il est nécessaire d'utiliser des outils adaptés à chaque type de données, tels que des plateformes de big data, des outils d'intégration des données et des logiciels de gestion des données. Ces outils permettent de centraliser, d'organiser et de préparer les données pour les analyses futures (Magnisalis, 2021).

3ème phase : La préparation

Dans leur article, M. Y. Santos et al. (2017) décrivent la phase de préparation des données comme l'intégration, le filtrage et l'enrichissement des données. Lors de cette phase, on va transformer les données brutes collectées en un format exploitable et utile pour les analyses futures.

Cette étape de préparation des données est essentielle pour transformer les données brutes collectées en un format exploitable et utile pour les analyses futures.

Intégration des données

L'intégration des données consiste à consolider toutes les données provenant de différentes sources internes et externes en un seul endroit avec une structure cohérente et homogène. Cette permet de faciliter l'accès, l'analyse et l'utilisation des données. Elle permet de fusionner des données de divers formats (structurés, semi-structurés et non structurés) en une base de données unifiée, permettant ainsi aux utilisateurs de poser des requêtes et d'obtenir des réponses à partir d'une source unique (Santos et al., 2017).

Filtrage des données

Le filtrage des données consiste à purifier les données en éliminant les éléments indésirables tels que les erreurs et les doublons. Ce processus inclut la détection et la correction des anomalies ainsi que l'anonymisation des informations sensibles pour garantir la confidentialité des données.

En nettoyant les données, le filtrage assure que seules des données de haute qualité sont utilisées dans les analyses, améliorant ainsi la fiabilité des résultats et des décisions basées sur ces analyses. Le résultat final est un ensemble de données catégorisées, purifiées et prêtes pour des analyses approfondies ou des recherches futures (Santos et al., 2017).

Enrichissement des données

L'enrichissement des données consiste à améliorer les données collectées en y ajoutant des informations contextuelles supplémentaires provenant de diverses sources. Ce processus vise à normaliser, simplifier et compléter les données brutes pour les rendre plus riches et informatives.

L'enrichissement inclut l'intégration de nouvelles données pertinentes, la correction des lacunes et l'ajout de contexte pour améliorer la qualité et la pertinence des données pour l'analyse.

En pratique, l'enrichissement des données peut impliquer l'ajout de données démographiques à un ensemble de données clients, l'intégration de données météorologiques à des données agricoles, ou l'ajout de données économiques à des rapports financiers. Le résultat est un ensemble de données raffinées et matures, prêtes pour des analyses approfondies ou pour être archivées à des fins de recherche future.

En enrichissant les données, les organisations peuvent obtenir des informations plus complètes et précises, ce qui améliore la prise de décision et permet de mieux comprendre les tendances et les comportements. L'enrichissement données sert donc à maximiser la valeur des données et optimiser les processus analytiques (Santos et al., 2017).

4ème phase : L'analyse

L'analyse de données est une phase clé du cycle de vie des données, au cours de laquelle une entreprise traite et interprète les ensembles de données à sa disposition pour en extraire des connaissances et des aperçus pertinents.

Selon C. Cote (2021), chaque analyse de données nécessite la définition de plusieurs paramètres liés aux indicateurs que l'on souhaite évaluer. Ces paramètres incluent la période sur laquelle l'analyse doit porter, la source des données et le type de données analysées.

Cette phase implique le développement de toutes les analyses et de tous les outils analytiques nécessaires pour extraire des connaissances et découvrir de nouvelles perspectives à partir des données.

Divers outils analytiques de big data sont utilisés pour analyser les données, permettant aux décideurs de comprendre ce qui se passe (analyse descriptive), pourquoi cela se passe (analyse causale), de simuler des scénarios hypothétiques (analyse what-if) et de faire des prévisions (analyse prédictive) :

- **Analyse descriptive** : Les techniques d'analyse descriptive permettent de synthétiser et de visualiser les données historiques pour identifier des tendances et des modèles.
- **Analyse causale** : L'analyse causale vise à déterminer les raisons sous-jacentes des phénomènes observés en explorant les relations entre les variables.
- **Analyse what-if** : Les scénarios hypothétiques (what-if) aident à évaluer les conséquences possibles de différentes décisions en simulant divers scénarios.
- **Analyse prédictive** : L'analyse prédictive utilise des modèles statistiques et des algorithmes de machine learning pour anticiper les événements futurs et informer la planification stratégique.

En combinant ces techniques, les entreprises peuvent obtenir une compréhension approfondie de leurs données, ce qui leur permet de prendre des décisions plus éclairées et de mieux anticiper les

défis et opportunités à venir. Cette approche analytique intégrée est essentielle pour maximiser la valeur des données et soutenir la prise de décision stratégique dans un environnement commercial de plus en plus complexe et concurrentiel (Cote, 2021).

5ème phase : La visualisation

La phase de visualisation concerne la présentation et l'interprétation des résultats des analyses de données. Cette étape est destinée aux utilisateurs finaux des entreprises, elle leur permet de comprendre facilement les informations découvertes et de prendre des décisions éclairées à l'aide de ces informations (Magnisalis, 2021).

Le National Institute of Standards and Technology (2015) détaille trois principales catégories de visualisation des données (paraphrasé par Magnisalis et al., 2021).

Visualisation exploratoire

La visualisation exploratoire vise à améliorer la compréhension des données, en particulier lorsque les ensembles de données sont volumineux. Elle inclut des techniques telles que la navigation dans les données, la détection des conditions limites et l'identification des valeurs aberrantes. Cette approche est particulièrement utile pour traiter de grandes quantités de données et nécessite souvent de nouvelles méthodes pour analyser efficacement ces volumes (paraphrasé Magnisalis et al., 2021).

Visualisation explicative

La visualisation explicative se concentre sur les résultats analytiques. Elle comprend des activités telles que la confirmation des résultats, l'interprétation des analyses et la présentation des résultats en temps quasi réel. Par exemple, elle permet de confirmer des hypothèses ou d'interpréter les résultats analytiques de manière claire et concise. Cette catégorie de visualisation est essentielle pour comprendre en profondeur les résultats obtenus et pour les partager de manière efficace avec les parties prenantes (paraphrasé par Magnisalis et al., 2021).

Visualisation narrative

La visualisation explicative a pour objectif de raconter une histoire et de présenter les résultats de manière simple et accessible pour le grand public. Des exemples courants incluent les rapports de business intelligence, les résumés et les tableaux de bord. Cette approche permet de transformer des données complexes en informations compréhensibles et actionnables, facilitant ainsi la prise de décision pour les non-spécialistes (paraphrasé par Magnisalis et al., 2021).

Les résultats de cette phase peuvent être présentés sous différentes formes telles que des tableaux de bord, des présentations orales, des interactions avec les utilisateurs ou des rapports. Chaque format a ses avantages spécifiques et peut être choisi en fonction des besoins de l'audience cible.

La fonction principale de cette phase est de rendre les données accessibles aux décideurs non techniques, en les aidant à comprendre et à utiliser les résultats pour une prise de décision efficace. La visualisation des données permet de simplifier des informations complexes et de les rendre plus digestes, ce qui est crucial pour maximiser l'impact des analyses et faciliter une prise de décision stratégique éclairée (paraphrasé par Magnisalis et al., 2021).

6ème et 7ème phases : L'accès et le partage

Les phases d'accès et de partage suivent directement la phase de visualisation et combinent les éléments essentiels pour garantir que les données sont accessibles de manière sécurisée et partagées efficacement avec les parties prenantes pertinentes.

Accès aux données

Cette étape implique la mise en place de mécanismes de contrôle d'accès et d'authentification pour sécuriser l'accès aux données. Il est vital de protéger les informations sensibles et de s'assurer que les restrictions d'accès sont respectées. Les données critiques doivent être stockées de manière à permettre une récupération rapide et efficace en cas de besoin. Cela inclut l'utilisation de systèmes de gestion des accès, de protocoles de sécurité robustes et de solutions de sauvegarde pour garantir l'intégrité et la disponibilité des données (Magnisalis et al., 2021).

Partage des données

Le partage des données préparées avec les parties prenantes internes et externes est une étape importante du cycle de vie des données. Les fournisseurs de données mettent en œuvre des outils modernes, tels que les technologies basées sur les API (Application Programming Interface), pour promouvoir un partage sécurisé et efficace. Ils identifient les données non classifiées à publier et établissent des accords de partage de données en conformité avec les spécifications éthiques et légales. Il est également essentiel de suivre les lignes directrices pour la publication des données ouvertes et de maintenir un équilibre entre la disponibilité et la redondance des données (Magnisalis et al., 2021).

8ème phase : L'utilisation et le feedback

Cette phase consiste en l'appropriation des données par les parties prenantes y ayant accès, leur permettant d'exploiter les informations précieuses pour adapter leurs politiques et stratégies en conséquence. Les utilisateurs analysent les données pour en tirer des insights pertinents et appliquent ces connaissances pour améliorer leurs processus et décisions.

En outre, les utilisateurs apportent un feedback aux fournisseurs de données sous forme de commentaires et de suggestions. Ce retour d'information est essentiel pour identifier les améliorations nécessaires et garantir que les données restent précises, pertinentes et utiles. Les fournisseurs de

données examinent ces feedbacks et publient des versions mises à jour des ensembles de données après avoir intégré les retours des utilisateurs (Magnisalis et al., 2021).

9ème phase : L'archivage

Selon Johnson et al. (2024), l'archivage des données une étape important dans le processus de la gestion des données en entreprise. Ce processus consiste à conserver de manière sécurisée des informations qui ne sont plus utilisées activement, mais qui pourraient être nécessaires à l'avenir. L'archivage permet de préserver l'intégrité des données, de gérer efficacement les coûts de stockage et de garantir la conformité aux régulations en vigueur.

Préservation historique : L'archivage des données permet de conserver des enregistrements historiques essentiels. Ces archives peuvent être utilisées par les entreprises pour des audits, des analyses de tendances ou des recherches historiques. En maintenant l'intégrité et la qualité des données archivées, les entreprises s'assurent que ces informations restent fiables pour une utilisation future.

Conformité réglementaire : De nombreuses industries sont soumises à des régulations strictes concernant la conservation et la protection des données. L'archivage des données avec des mesures de sécurité adéquates garantit que les entreprises respectent ces régulations, évitant ainsi les risques de sanctions et protégeant leur réputation.

Optimisation des coûts : Les données actives nécessitent souvent des solutions de stockage coûteuses. L'archivage permet de transférer les données non actives vers des solutions de stockage plus économiques, optimisant ainsi les coûts liés au stockage tout en conservant l'accessibilité des informations en cas de besoin.

Sécurité des données : En mettant en place des mesures de sécurité robustes, telles que le chiffrement et les contrôles d'accès, les entreprises peuvent protéger les informations sensibles contre les accès non autorisés et les fuites. Cela est particulièrement important pour les entreprises qui doivent se conformer à des régulations strictes.

Accessibilité future : Les données archivées, bien que non actives, restent accessibles pour des analyses futures. Cela permet aux entreprises d'utiliser leurs archives pour des analyses de tendances, des prises de décisions stratégiques et des recherches approfondies, garantissant que les données restent un atout précieux pour l'organisation à long terme.

L'archivage des données doit être réalisé selon un plan formel, définissant les critères de conservation et les stratégies de récupération en cas de besoin. De plus, il est essentiel de sécuriser rigoureusement l'accès aux données archivées pour protéger ces informations contre toute violation de sécurité ou tout accès non autorisé (Johnson et al., 2024).

10ème phase : La fin de vie

La dernière phase d'une donnée au sein du cycle de vie des données est celle de sa fin de vie. L. Lin et al. (2014) décrivent cette étape comme l'action consistant à éliminer du système les données dupliquées, non requises ou inutiles. Cette phase est essentielle pour gérer efficacement les ressources et éviter l'accumulation de données obsolètes (Lin et al., 2014).

Les données définies comme étant en fin de vie doivent être détruites conformément aux règles et régulations en vigueur afin d'assurer une suppression sécurisée. Cela garantit que toutes les données non nécessaires sont correctement éliminées.

Lors de cette phase, il est important de s'assurer que les données non nécessaires sont supprimées de façon permanente et ne peuvent pas être restaurées à partir du support de stockage. Cela prévient la divulgation accidentelle d'informations sensibles (Lin et al., 2014).

Cette dernière étape marque la fin du parcours des données au sein d'une entreprise.

Cependant, il existe quatre phases additionnelles dans ce cycle de vie, qui sont des phases transversales. C'est-à-dire qu'elles se déroulent parallèlement à toutes les autres étapes du cycle des données. Ces quatre phases sont la qualité des données, le stockage des données, la protection et la sécurité des données et la gouvernance des données.

1.3.2 Phases transversales du cycle de vie des données

La qualité des données

Selon F.J. Scheuren et al. (2007), la qualité des données représente leur "aptitude à l'utilisation", autrement dit, la capacité des données à remplir leurs rôles prévus au sein de l'organisation et à être conformes à des normes préétablies.

Dans l'article " Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality " (2023), J. Wang et al. évaluent cette aptitude à l'utilisation à travers plusieurs dimensions, qui se regroupent en trois grandes catégories : la qualité intrinsèque, la qualité contextuelle et la qualité représentative. Chacune de ces catégories englobe des dimensions spécifiques qui contribuent à une compréhension globale et approfondie de la qualité des données.

La **qualité intrinsèque** des données se concentre sur les caractéristiques inhérentes aux données elles-mêmes, indépendamment de leur contexte d'utilisation. Elle inclut les dimensions suivantes :

- **Précision (ou exactitude)** : Les données doivent refléter avec exactitude la réalité ou les valeurs correctes, sans erreur.
- **Crédibilité** : Les données doivent être perçues comme fiables et dignes de confiance par les utilisateurs.
- **Objectivité** : Les données doivent être impartiales et exemptes de biais.
- **Réputation** : La source des données doit être reconnue et respectée pour sa fiabilité.

Ces dimensions garantissent que les données sont correctes, fiables et proviennent de sources dignes de confiance.

La **qualité contextuelle** des données prend en compte la pertinence et l'adéquation des données pour des tâches spécifiques. Cette catégorie est composée des dimensions suivantes :

- **Complétude** : Toutes les données nécessaires doivent être présentes pour répondre aux besoins spécifiques.
- **Pertinence** : Les données doivent être pertinentes pour le contexte particulier dans lequel elles sont utilisées.
- **Rapidité (ou actualité)** : Les données doivent être à jour et disponibles au moment requis pour une prise de décision efficace.
- **Volume approprié** : La quantité de données doit être suffisante pour répondre aux besoins contextuels sans être excessive.

Ces dimensions assurent que les données sont appropriées et suffisantes pour le contexte spécifique dans lequel elles sont utilisées.

La **qualité représentative** des données concerne la manière dont les données sont présentées et interprétées. Elle comprend les dimensions suivantes :

- **Représentation concise** : Les données doivent être représentées de manière claire et concise, facilitant leur compréhension.
- **Facilité de manipulation** : Les données doivent être faciles à manipuler et à analyser, sans nécessiter de transformations complexes.
- **Interprétabilité** : Les données doivent être compréhensibles et faciles à interpréter pour les utilisateurs.
- **Cohérence** : Les données doivent être uniformes et ne pas présenter de contradictions entre différentes sources ou périodes.

Ces dimensions garantissent que les données sont présentées de manière claire et cohérente, facilitant leur utilisation et leur interprétation par les utilisateurs (Wang et al., 2023).

Toujours selon J. Wang et al. (2023), pour assurer la qualité des données, il existe plusieurs méthodes qu'il est bon de combiner pour avoir le meilleur résultat possible.

En premier lieu, on va avoir des **normes de qualité** de données qui incluent des critères tels que la précision, la complétude, la cohérence, la validité et la pertinence. Ces normes serviront de référence pour toutes les opérations de gestion des données

On peut également mettre en place des **processus** rigoureux de **validation** et de **vérification** pour s'assurer que les données soient correctes et complètes avant leur utilisation. Cela peut inclure des contrôles automatisés ainsi que des revues manuelles

Il est également important de **former** les **employés** sur l'importance de la qualité des données et sur les meilleures pratiques pour la gérer. Des sessions de formation régulières et des programmes de sensibilisation peuvent aider à maintenir un haut niveau de qualité des données

Avoir des données de qualité a de nombreux impacts positifs sur une entreprise. Selon G. Georgiadis et G. Poels (2021), des données complètes et pertinentes permettent une prise de décision éclairée grâce à des analyses fiables. De plus, des données actualisées et pertinentes améliorent l'efficacité opérationnelle de l'entreprise. La satisfaction client est également influencée par la qualité des données, car des informations à jour permettent d'anticiper les besoins des clients et d'y répondre plus efficacement. Enfin, un avantage concurrentiel significatif est obtenu lorsque les données sont faciles à manipuler et à interpréter, permettant à l'entreprise de réagir rapidement aux changements du marché et de saisir de nouvelles opportunités.

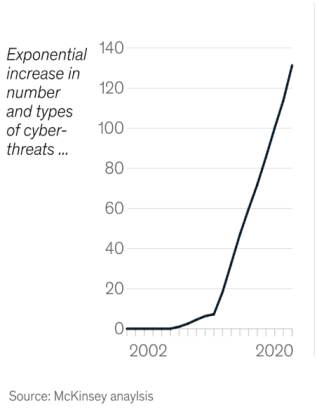
La sécurité des données

La sécurité des données est un ensemble de pratiques et de technologies destinées à protéger les informations sensibles contre les accès non autorisés, les altérations et la destruction. Elle garantit la confidentialité, l'intégrité et la disponibilité des données. Dans le contexte actuel, la sécurité des données est devenue un enjeu majeur pour les entreprises (Magnisalis et al., 2021).

Comme le montre la Figure 4, on assiste à une augmentation exponentielle des cyberattaques depuis plusieurs années, ces attaques étant de plus en plus sophistiquées. Les menaces telles que les attaques par ransomware et le phishing représentent des risques majeurs pour les entreprises. Les conséquences des violations de données peuvent être dévastatrices, entraînant des pertes financières, des atteintes à la réputation et des sanctions légales (Boehm et al. 2022).

Les entreprises doivent également se conformer à des exigences en matière de sécurité de plus en plus strictes. Par exemple, le Règlement Général sur la Protection des Données (RGPD) en Europe impose des normes rigoureuses pour la gestion et la protection des données personnelles. La conformité à ces régulations est essentielle pour éviter des amendes lourdes et préserver la confiance des clients. Avoir une bonne protection des données permet non seulement de se conformer aux lois, mais aussi de maintenir la confiance des clients et la réputation de l'entreprise (Magnisalis et al., 2021).

Figure 4 : Nombre de logiciels malveillants uniques recensés par année (en millions)



Source : McKinsey & Company (2022) Cybersecurity blanket https://www.mckinsey.com/featured-insights/sustainable-inclusive-growth/chart-of-the-day/cybersecurity-blanket

Le RGPD en Europe

Le Règlement Général sur la Protection des Données (RGPD) est une loi sur la confidentialité et la sécurité des données conçue par l'Union Européenne. Bien que ce règlement ait été rédigé et adopté par l'Union Européenne, il impose des obligations aux organisations du monde entier, à condition qu'elles ciblent ou collectent des données relatives aux personnes situées dans l'UE. Le RGPD est entré en vigueur le 25 mai 2018 et prévoit des sanctions sévères pour les contrevenants, pouvant atteindre des dizaines de millions d'euros (Wolford, s. d.).

Le RGPD a été mis en place pour renforcer la protection de la vie privée des individus dans un contexte où de plus en plus de données personnelles sont partagées en ligne et où les violations de données sont fréquentes (Wolford, s. d.).

Parmi les points les plus importants du RGPD, on trouve :

1. **Consentement** : Les entreprises doivent obtenir un consentement explicite, informé et non ambigu des individus avant de collecter leurs données.
2. **Droits des individus** : Les citoyens européens disposent de droits renforcés sur leurs données, notamment le droit d'accès, de rectification, d'effacement, et de portabilité de leurs données.
3. **Notification de violation des données** : En cas de violation de données, les entreprises doivent notifier les autorités compétentes et les personnes concernées dans un délai de 72 heures.
4. **Protection des données dès la conception** : Toute nouvelle activité ou produit doit intégrer la protection des données dès la conception et par défaut.
5. **Responsabilité des entreprises** : Les entreprises doivent prouver leur conformité au RGPD, en maintenant une documentation détaillée et en mettant en place des mesures techniques et organisationnelles appropriées pour sécuriser les données.

Application du RGPD par les entreprises

Pour se conformer au RGPD, les entreprises doivent adopter des politiques de gestion des données robustes et des mesures de sécurité avancées. L'entreprise de cybersécurité Vistrada a cité certains de ces processus dans son article de décembre 2023 " Data security compliance: standards, regulations, and best practices".

En premier, on peut citer le fait pour une entreprise de faire des audits réguliers pour évaluer les risques et identifier les vulnérabilités au sein de son infrastructure et mettre en œuvre des mesures correctives. Les audits doivent inclure des évaluations de la conformité au RGPD et à d'autres réglementations pertinentes.

Il est également important d'utiliser des technologies de chiffrement pour protéger les données sensibles, tant en transit (lorsqu'elles sont transférées d'un point à un autre, comme via un réseau ou Internet) qu'au repos (lorsqu'elles sont stockées sur des serveurs, des bases de données ou des disques durs). Ces mesures de sécurité sont vitales pour garantir que les données restent confidentielles et protégées contre tout accès non autorisé, qu'elles soient en cours de transfert ou stockées de manière statique.

Le chiffrement garantit que seules les personnes autorisées, disposant de la clé de déchiffrement appropriée, peuvent lire les données. Il est également conseillé de séparer des données jugées comme sensibles (données médicales ou bancaires par exemple) des autres données au sein des services cloud.

Vistrada recommande également d'implémenter des politiques de contrôle d'accès strictes basées sur le principe du moindre privilège, où les utilisateurs n'ont accès qu'aux données nécessaires à l'exécution de leurs tâches.

La surveillance et la détection des menaces jouent un rôle clé dans la sécurité des données. En utilisant des outils de surveillance et d'analyse comportementale, les organisations peuvent identifier et répondre rapidement aux incidents de sécurité en détectant les activités suspectes.

La sensibilisation et la formation des employés sont également fondamentales. Il est important de former régulièrement les employés sur les meilleures pratiques en matière de sécurité des données et sur les procédures de réponse aux incidents. Les employés doivent être conscients des menaces potentielles et savoir comment les éviter.

Enfin, la sécurité des données implique de développer et de tester des plans de réponse aux incidents pour gérer efficacement les violations de données et en limiter les impacts. Ces plans doivent inclure des procédures de notification conformes au RGPD, afin d'assurer une gestion appropriée et légale des incidents de sécurité.

Une bonne protection des données permet à une entreprise de s'éviter des problèmes qui peuvent s'avérer fatals, comme une amende pouvant monter jusqu'à 20 millions d'euros pour non-respect du RGDP ou encore une perte de confiance des clients envers une entreprise qui aurait subi une fuite de données sensibles (Vistrada, 2023).

Le stockage des données

Le stockage des données se réfère à la méthode par laquelle les informations numériques sont conservées dans des systèmes informatiques. Cette phase est essentielle tout au long du cycle de vie des données, car elle permet de sauvegarder les informations pour un accès futur, la récupération, l'analyse et le traitement des données (Magnisalis et al., 2021).

Il existe différents supports digitaux pour stocker les données, je vais en développer certains ici.

- **Disques durs (HDD ou SSD)** : Les disques durs mécaniques (Hard Disk Drive) utilisent des plateaux rotatifs et un bras mécanique pour lire et écrire des données, offrant une grande capacité de stockage à faible coût, mais avec des vitesses plus lentes en raison de la nature mécanique du processus. En revanche, les disques à état solide (Solid State Drive) utilisent des puces de mémoire flash sans pièces mobiles, permettant des temps de lecture/écriture beaucoup plus rapides et une meilleure durabilité. Les SSD consomment également moins d'énergie et génèrent moins de chaleur, ce qui en fait un choix supérieur pour les performances globales, tandis que les HDD sont préférés pour leur capacité économique à stocker de grandes quantités de données (Crucial, 2014).
- **Stockage en réseau** : Un système de stockage de réseau ou NAS (Network Attached Storage) est un dispositif de stockage de données connecté à un réseau informatique, permettant aux utilisateurs du réseau d'accéder et de partager des fichiers à partir d'un emplacement centralisé. Un NAS fonctionne comme un serveur dédié, conçu spécifiquement pour le stockage de fichiers, offrant des services de stockage et de partage de fichiers à d'autres appareils connectés au réseau. Cette solution permet une gestion simplifiée du stockage grâce à une interface centralisée et un accès sécurisé par différentes personnes depuis différents appareils (Mistry et al., 2020).
- **Stockage en cloud** : Le stockage en cloud permet de stocker des données sur des serveurs distants gérés par des fournisseurs de services tels que Amazon ou Google. Ce modèle offre aux utilisateurs la flexibilité d'ajuster la capacité de stockage en fonction des besoins, sans nécessiter d'investissement dans des infrastructures matérielles. Les utilisateurs peuvent accéder à leurs données via Internet depuis n'importe quel appareil, facilitant le travail à distance et la collaboration (Joshi, 2024).

Importance du stockage

Le stockage des données représente une composante importante dans les opérations des entreprises modernes, avec des impacts significatifs sur plusieurs aspects clés (Opsmatters, 2024) :

- **Accessibilité de l'information** : Un système de stockage de données bien conçu permet un accès rapide et facile aux informations critiques, ce qui est essentiel pour la prise de décisions rapides et la collaboration efficace au sein des équipes. Cela facilite la réponse aux demandes des clients et soutient la productivité générale.
- **Analyse des données et perspectives** : Le stockage de grandes quantités de données permet aux entreprises de mener des analyses approfondies, offrant des perspectives précieuses sur

les tendances, les comportements des clients, et les opportunités de marché. Ces informations aident à optimiser les opérations, améliorer l'expérience client, et stimuler l'innovation.

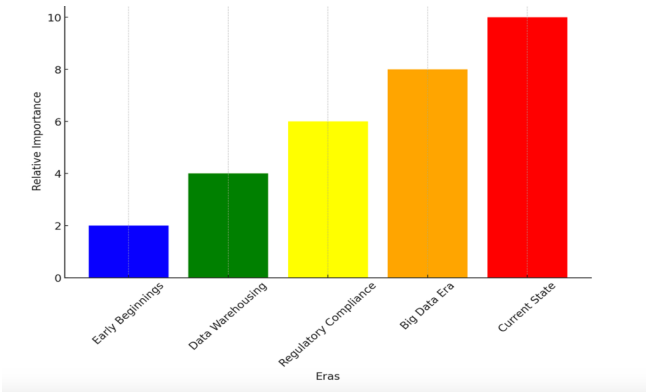
- **Amélioration de la prise de décisions** : Le stockage et l'analyse des données permettent aux entreprises de baser leurs décisions sur des preuves concrètes plutôt que sur des intuitions. Cela conduit à des stratégies plus efficaces et à des résultats optimisés.
- **Conformité réglementaire** : Le respect des réglementations sur la protection des données, telles que le RGPD, est crucial pour éviter des sanctions et protéger la réputation de l'entreprise. Un système de stockage conforme aux normes réglementaires garantit que les données sont gérées en toute sécurité, avec des mesures comme le chiffrement et les contrôles d'accès en place.

1.4 Gouvernance des données

Sur le schéma DaLiF (Figure 3), les 3 phases transversales du cycle de vie des données que sont la qualité des données, le stockage des données et la protection et sécurité des données sont liées à une autre phase transversale du cycle de vie des données que l'on appelle la gouvernance des données.

La gouvernance des données est un concept dont la signification a beaucoup évolué au fil des années, pour répondre aux besoins croissants des organisations en matière de gestion des données. Comme indiqué sur la Figure 5, D. Sargiotis (2024) a divisé cette évolution en 5 périodes distinctes, liées à des évolutions techniques et réglementaires.

Figure 5 : Évolution historique de l'importance de la gouvernance des données



Source : Sargiotis D. (2024) Data Governance in the Digital Age: Strategies, Challenges, and Best Practices

À l'origine, la gestion des données était principalement une préoccupation technique, axée sur le stockage et la maintenance des données. Bien que le concept de gouvernance des données ne soit pas encore formalisé, les organisations commençaient à reconnaître l'importance de l'exactitude et de la sécurité des données (Sargiotis, 2024).

Avec l'avènement des entrepôts de données dans les années 1990, les données ont commencé à être perçues comme une ressource précieuse pour l'intelligence d'affaires. Cette période a marqué le début des pratiques systématiques de gestion des données, bien que celles-ci soient encore principalement dirigées par les services informatiques (Sargiotis, 2024).

Le début des années 2000 a vu une montée en puissance des exigences réglementaires, comme la loi Sarbanes-Oxley de 2002 aux États-Unis, qui a mis en lumière la nécessité d'une gouvernance formelle des données pour assurer la conformité et gérer les risques associés. Les organisations ont commencé

à comprendre l'importance de la gouvernance des données pour répondre aux exigences réglementaires et réduire les risques (Sargiotis, 2024).

L'explosion du big data dans les années 2010, alimentée par l'avènement des réseaux sociaux, de l'Internet des objets (IoT) et d'autres technologies numériques, a apporté de nouveaux défis et opportunités. La gouvernance des données s'est alors étendue pour inclure des aspects tels que la confidentialité des données, l'utilisation éthique des données, et la nécessité de techniques plus sophistiquées de qualité et d'intégration des données (Sargiotis, 2024).

Aujourd'hui on définit la gouvernance des données comme "l'ensemble de processus, de politiques, de normes et de mesures qui garantissent l'utilisation efficace et efficiente de l'information, permettant ainsi à une organisation d'atteindre ses objectifs" (Sargiotis, 2024). Ce concept vise à optimiser l'efficacité des informations disponibles grâce à des processus techniques performants et à un engagement organisationnel global.

Cette définition souligne l'objectif de la gouvernance des données : optimiser l'efficacité des informations disponibles. Cela implique non seulement des processus techniques performants, mais aussi un engagement organisationnel global. Une approche holistique de la gouvernance des données permet de gérer l'exactitude, la cohérence, l'accessibilité et la sécurité des données tout au long de leur cycle de vie (Sargiotis, 2024).

Comme l'a montré le contexte théorique, les données ont pris une importance toujours plus capitale au sein des entreprises, en raison du volume exponentiel de données à traiter, des nouvelles réglementations à respecter et des possibilités d'informations qu'offrent les nouvelles technologies. Ce contexte pousse de plus en plus d'entreprises à mettre en place une gouvernance des données pour répondre efficacement aux défis posés par cet engouement autour des données (Sargiotis, 2024).

Apports d'une bonne gouvernance des données

Une gouvernance des données ne va pas produire de revenus directs à l'entreprise, ou une baisse de coûts ou de risques. Cependant, implémenter une bonne gouvernance des données apporte une multitude d'avantages dans divers aspects des opérations d'une entreprise qui vont eux affecter directement la performance de l'entreprise. On peut citer plusieurs aspects qui seront impactés positivement par une bonne gouvernance des données (Data Governance Institute, s.d.) :

- **Améliorer la prise de décisions** : En garantissant que les données sont précises, complètes et disponibles en temps opportun, les décisions prises par l'organisation sont mieux informées et plus fiables
- **Réduire les frictions opérationnelles :** En standardisant les processus et en clarifiant les responsabilités, la gouvernance des données aide à minimiser les inefficacités et les conflits internes
- **Protéger les besoins des parties prenantes des données** : Veiller à ce que chaque groupe ou individu qui utilise ou est affecté par les données de l'organisation ait ses besoins et préoccupations pris en compte et traités de manière appropriée

- **Former le management et le personnel :** Encourager l'adoption de méthodes communes pour gérer les problèmes liés aux données, renforçant ainsi la cohérence et l'efficacité
- **Établir des processus standardisés et répétables :** Mettre en place des processus clairs et uniformes pour la gestion des données, ce qui facilite la répétition et la fiabilité des opérations
- **Réduire les coûts et augmenter l'efficacité :** Coordonner les efforts pour éviter les duplications et optimiser l'utilisation des ressources.
- **Assurer la transparence des processus :** Garantir que les processus de gouvernance des données sont transparents et compréhensibles pour toutes les parties prenantes.

Ces différents aspects vont se retrouver au sein d'une entreprise qui a mis en place un cadre de gouvernance des données efficace, et ainsi permettre à une entreprise de vraiment appréhender ses données comme leviers stratégiques, stimulant la croissance et l'innovation tout en bâtissant une relation de confiance durable avec les clients.

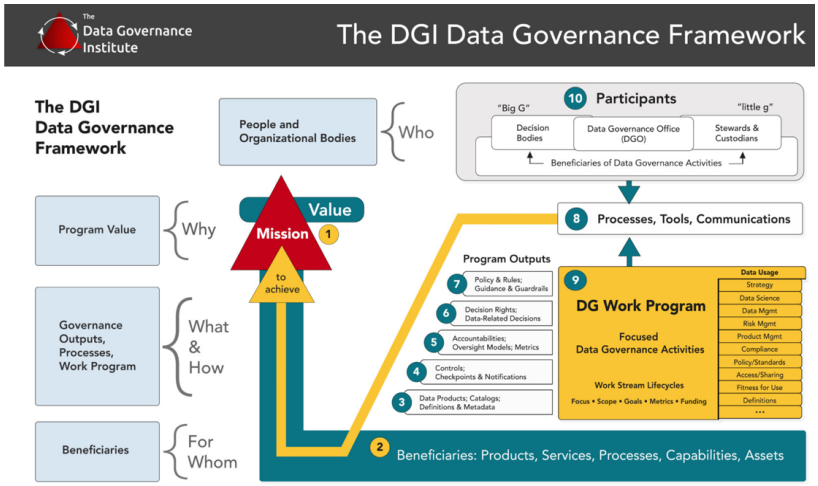
En pratique, selon un récent rapport de l'université de Drexel (2023) qui a interrogé 450 experts des données dans le monde, 57% des entreprises qui ont mis en place un programme de gouvernance des données voient une amélioration dans la qualité des analyses des données et des informations tirées, et 60% des experts data voient une amélioration dans la qualité de la donnée elle-même.

1.4.1 Mise en œuvre d'un cadre de gouvernance des données

Pour mettre en œuvre un cadre de gouvernance des données, je me base dans ce mémoire sur celui construit par le Data Governance Institute (DGI), le Data Governance Framework (Figure 6). Cette organisation a été fondée en 2003 par Gwen Thomas pour répondre à un besoin croissant de structuration et d'organisation des activités complexes liées à la gestion des données d'entreprise. C'est dans cette idée qu'elle a développé, et mis régulièrement à jour, un modèle de gouvernance des données, servant de modèle à de nombreuses entreprises à travers le monde (Data Governance Institute, s. d.).

Je vais décrire la version la plus récente de ce modèle afin d'avoir une base sur laquelle proposer un prototype qui serait adapté à ce qu'on veut développer au sein de Peppl.

Figure 6 : Le Data Governance Framework



Source : The Data Governance Institute (s.d.) The Data Governance Framework

Ce modèle de gouvernance des données se divise en 3 parties principales :

1. Mission et valeurs ajoutées par la gouvernance des données
2. Composantes d'une gouvernance des données
3. Participants à la gouvernance des données

1.4.2 Mission et valeurs ajoutées par la gouvernance des données

Avant la mise en place d'une gouvernance des données, le DGI (s.d.) estime qu'il est nécessaire de définir les raisons pour lesquelles un tel programme est nécessaire pour l'organisation et les objectifs que l'on souhaite atteindre. Pour cela, il est essentiel de répondre à diverses questions concernant la gestion initiale des données, la perception managériale, les risques de sécurité, ainsi que les attentes spécifiques des différents départements. Ces questions permettent de clarifier les priorités et de définir une mission claire pour la gouvernance des données, une mission qui doit impérativement apporter de la valeur à l'entreprise.

Un programme de gouvernance des données bien conçu peut offrir de la valeur sous plusieurs formes. Par exemple, il peut contribuer à augmenter les revenus en permettant la création de produits de données monétisables ou en fournissant des informations précieuses pour stimuler les ventes. En améliorant la qualité et la fiabilité des données, il peut également réduire les coûts, en particulier ceux liés au stockage et à la gestion des données redondantes ou erronées, et minimiser les erreurs dans le développement de logiciels.

En outre, un cadre de gouvernance des données axé sur la conformité réglementaire et la sécurité peut aider l'entreprise à éviter les sanctions et les dommages à sa réputation en cas de violation des données.

Il faut également que la gouvernance des données soit alignée avec les objectifs commerciaux de l'entreprise, qu'il s'agisse d'améliorer l'efficacité opérationnelle, d'accroître la satisfaction des clients ou de favoriser l'innovation.

Le programme de gouvernance doit être adapté aux spécificités de chaque entreprise, en tenant compte de son contexte, de ses activités, et des parties prenantes concernées. Ce programme peut occuper une place plus ou moins centrale dans l'organisation, allant d'une fonction stratégique nécessitant l'engagement de l'ensemble de l'entreprise à une fonction plus ciblée, apportant de la valeur à des départements ou individus spécifiques.

En somme, la clé du succès d'un programme de gouvernance des données réside dans sa capacité à soutenir directement les objectifs commerciaux de l'entreprise tout en s'adaptant à ses besoins et ressources uniques (Data Governance Institute, s. d.).

1.4.3 Composantes d'une gouvernance des données

Un programme de gouvernance des données comprend plusieurs composantes clés.

Politiques et règles des données

Selon le DGI (s.d.), un programme de gouvernance des données bien établi contribue significativement à la création de politiques de données descendantes (top-down) dans l'entreprise. Une politique d'entreprise, ce sont les lignes directrices à suivre formulées à l'attention des membres de l'entreprise pour pouvoir appliquer des procédures définies. Cela signifie que mettre en place un programme de gouvernance des données nécessite la définition de règles et de normes à un niveau stratégique, influençant ainsi l'orientation globale de l'organisation en matière de gestion des données.

Ces politiques vont agir comme des “traducteurs” entre les différentes équipes au sein de l'organisation. Ils interprètent les besoins et les contraintes des entreprises en termes de produits, services, processus, capacités et actifs, puis traduisent ces besoins en termes techniques compréhensibles pour les équipes juridiques et techniques. Cela permet de garantir que toutes les parties prenantes comprennent les implications des politiques de données et peuvent agir en conséquence (Data Governance Institute, s. d.).

En apportant une prise de conscience des changements, des opportunités ou des préoccupations, les équipes de gouvernance des données déclenchent des analyses et des collaborations. Cela permet de s'assurer que les politiques de données sont non seulement définies, mais aussi mises en œuvre de manière efficace et cohérente à travers l'organisation (Petzold et al., 2020).

Droits de décision et responsabilités

Prises de décisions

Un programme de gouvernance des données basé sur celui du DGI a pour mission de faciliter, documenter, stocker et rendre disponibles les informations sur les décisions prises.

Cela inclut la mise en place de processus clairs pour la prise de décisions, l'enregistrement de chaque décision en précisant l'auteur, la date et la raison, ainsi que la conservation de ces enregistrements de manière accessible pour des références futures.

Les décisions concernant la conformité aux lois, réglementations ou normes sont généralement prises au niveau exécutif. Toutefois, la manière de se conformer peut nécessiter la contribution de plusieurs parties prenantes pour évaluer les différentes options disponibles. Pour d'autres types de décisions, il est essentiel d'identifier les produits, services et actifs concernés, puis de choisir des représentants en charge de la prise de décision, garantissant ainsi que toutes les parties prenantes sont considérées et que leurs besoins sont pris en compte dans le processus décisionnel.

Un bon programme de gouvernance des données permet de documenter les décisions prises et les participants impliqués, ce qui est utile pour la transparence et l'évaluation des processus décisionnels. Cela aide à vérifier si les décisions ont été prises en utilisant les bons critères et le bon contexte (Data Governance Institute, s. d.).

Responsabilités

Après avoir établi qui a le rôle de décisionnaire, le cadre de gouvernance des données du DGI permet également de définir les responsabilités, en identifiant les personnes responsables de chaque tâche et en déterminant les délais et les calendriers. Pour les activités ne correspondant pas directement aux responsabilités d'un département spécifique, le programme peut clarifier qui est responsable de chaque tâche et intégrer ces responsabilités dans les processus quotidiens, assurant ainsi qu'elles fassent partie des opérations normales.

Plutôt que de laisser chaque gestionnaire interpréter indépendamment les exigences de conformité, les organisations centralisent souvent le développement des exigences, créées par un groupe central et communiquées à tous les départements concernés.

En résumé, une fois qu'une règle est créée ou qu'une décision liée aux données est prise, l'un des enjeux est de clarifier les responsabilités et de les intégrer dans les processus quotidiens et les cycles de développement de l'organisation. Le programme de gouvernance des données joue un rôle central dans cette tâche, en particulier pour les activités de conformité qui nécessitent une collaboration interfonctionnelle et une documentation rigoureuse (Data Governance Institute, s. d.).

Contrôles des risques

Avec la prolifération des violations de données et les conséquences qui en découlent, il est évident que les données peuvent représenter un risque significatif. Pour gérer ce risque, il est essentiel d'empêcher les événements indésirables de se produire et, pour ceux qui ne peuvent être complètement évités, de les détecter rapidement afin de pouvoir corriger les problèmes (Data Governance Institute, s. d.).

À travers un programme de gouvernance des données, on définit les contrôles nécessaires pour atténuer ces risques. Ces contrôles peuvent être classés en deux catégories principales : les contrôles préventifs, qui visent à empêcher les incidents avant qu'ils ne se produisent, et les contrôles détectifs et correctifs, qui visent à identifier et à rectifier rapidement les incidents lorsqu'ils surviennent. En combinant ces approches, le programme de gouvernance des données assure une gestion proactive et réactive des risques, protégeant ainsi l'intégrité et la sécurité des données de l'organisation.

Les contrôles préventifs visent à empêcher les incidents avant qu'ils ne se produisent (Data Governance Institute, s. d.). Voici quelques exemples :

- **Politiques de sécurité des données** : Établir des politiques claires sur la gestion et la protection des données. Cela inclut des directives sur l'accès aux données, le cryptage des données sensibles et les protocoles de sauvegarde.
- **Formation et sensibilisation** : Former les employés sur les bonnes pratiques de gestion des données et les sensibiliser aux risques liés aux données. Une formation régulière permet de maintenir un niveau de vigilance élevé.
- **Contrôles d'accès** : Limiter l'accès aux données en fonction des rôles et des responsabilités des employés. Utiliser des technologies d'authentification et d'autorisation pour garantir que seules les personnes autorisées peuvent accéder aux données sensibles.

Les contrôles détectifs et correctifs permettent d'identifier les incidents lorsqu'ils se produisent et de les corriger rapidement (Data Governance Institute, s. d.). Voici quelques exemples :

- **Surveillance et audit** : Mettre en place des systèmes de surveillance pour détecter les accès non autorisés ou les anomalies dans l'utilisation des données. Effectuer des audits réguliers pour vérifier la conformité aux politiques de sécurité.
- **Alertes et notifications** : Configurer des alertes pour notifier les administrateurs en cas d'activités suspectes ou de violations potentielles. Cela permet de réagir rapidement pour contenir et corriger les incidents.
- **Plans de réponse aux incidents** : Développer des plans détaillés pour répondre aux incidents de sécurité des données. Cela inclut des procédures pour isoler les systèmes affectés, enquêter sur les incidents et restaurer les données à partir des sauvegardes.

Produits de données

Selon le DGI (s. d.), un programme de gouvernance des données développe généralement un ensemble de produits de données qui sont des actifs réutilisables et spécifiquement adaptés à différents usages.

Ces produits comprennent souvent des ensembles de données fiables, collectés à partir de diverses sources pertinentes et conçus pour garantir confidentialité, qualité et standardisation des données. En traitant les données brutes, ces produits assurent une gestion appropriée de la confidentialité et de la vie privée, tout en maintenant des normes élevées de qualité.

Ces produits de données démontrent l'efficacité du programme de gouvernance des données en fournissant des outils pratiques et utiles pour les membres de l'organisation. Souvent, ils sont accompagnés de certifications attestant qu'ils respectent les exigences de gouvernance et de conformité.

Parmi ces produits, on trouve des tableaux de bord, des ensembles de données structurés et des modèles de données, tous caractérisés par diverses informations telles que la période couverte, des instructions d'utilisation, les contraintes de partage et les exigences de conformité applicables.

En fournissant ces produits de données, le programme de gouvernance des données permet aux différents membres de l'entreprise d'accéder à des informations fiables et standardisées, facilitant ainsi la prise de décision et l'optimisation des processus internes (Data Governance Institute, s. d.).

Processus et outils utilisés

Au sein du programme de gouvernance des données du DGI, on va définir un ensemble de processus permettant la réalisation des tâches associées à ce programme.

Un processus métier est une série de tâches standardisées qu'une entreprise utilise pour atteindre des objectifs spécifiques de manière cohérente et efficace. Ces processus sont essentiels pour structurer et organiser les activités répétitives. Ils doivent répondre à trois critères (Villanova University, 2022) :

- **Répétabilité** : Garantir que le processus peut être exécuté de manière uniforme à chaque itération.
- **Transparence** : Permettre le suivi et l'évaluation des performances.
- **Agilité** : S’assurer que le processus puisse s'adapter facilement aux changements tout en restant efficace et pertinent.

Ces processus couvrent divers aspects, notamment l'alignement des politiques, l'établissement de responsabilités, la gestion du changement, et la définition de la qualité des données.

La mise en œuvre efficace de ces processus de gouvernance nécessite souvent l'utilisation d'outils technologiques (Data Governance Institute, s. d.).

Les outils technologiques peuvent jouer plusieurs rôles clés :

- **Gestion et suivi du programme** : Ils aident à orchestrer et à surveiller les activités du programme, permettant ainsi un suivi rigoureux des tâches et des résultats obtenus.
- **Engagement des participants** : Les plateformes collaboratives facilitent la participation active de toutes les parties prenantes, assurant une communication fluide et une adhésion aux processus.
- **Capture des métriques et évaluation** : Les outils de suivi de performance permettent de mesurer les performances du programme, fournissant des données concrètes pour évaluer l'efficacité des processus et pour ajuster les stratégies en conséquence.
- **Documentation et partage des décisions** : Ils permettent de centraliser la documentation des politiques, des règles, et des décisions, garantissant que tous les membres de l'équipe ont accès aux informations actualisées et nécessaires pour mener à bien leurs tâches.

Organisation des flux de travail

La dernière composante d'un programme de gouvernance des données suivant le modèle du DGI, c'est l'organisation et la gestion des flux de travail.

Un flux de travail est l'organisation systématique de processus menés par une personne ou une organisation de personnes visant à accomplir une tâche (van Zandt, 2023).

Dans le modèle de gouvernance des données du DGI, chaque flux de travail possède son propre cycle de vie, avec une portée, des objectifs, des métriques, un financement et des responsabilités spécifiques. Il est nécessaire de définir ces éléments clairement pour chaque flux de travail afin d'assurer une gestion efficace et une réalisation des objectifs

Les flux de travail peuvent produire des résultats variés. Certains génèrent des résultats anecdotiques ou difficiles à mesurer, tandis que d'autres doivent répondre aux critères SMART (Spécifiques, Mesurables, Actionnables, Pertinents, et Opportuns). Il est important de distinguer ces types de résultats pour évaluer correctement leur impact au sein de l'entreprise.

Pour chaque flux de travail, tous les participants à la gouvernance des données doivent comprendre à quoi ressemble le succès et comment il est mesuré. Il est recommandé de créer des déclarations de valeur en utilisant une formule précise pour clarifier les objectifs et les résultats attendus. Ces déclarations aident à aligner les efforts des participants et à fournir une direction claire pour chaque flux de travail (Data Governance Institute, s. d.).

1.4.4 Participants à la gouvernance des données

Pour mettre en place un cadre de gouvernance des données, les membres de l'entreprise doivent remplir des rôles spécifiques. Le modèle de gouvernance des données du DGI en définit plusieurs.

Acteurs de données et organes décisionnels

Les acteurs de données sont des individus ou des groupes qui influencent ou sont influencés par les décisions liées aux données. En raison de leur rôle central dans la gestion des données, ils ont généralement des attentes spécifiques qui doivent être prises en compte par le programme de gouvernance. Certains de ces acteurs participent directement à la prise de décision, tandis que d'autres sont consultés avant que les décisions ne soient formalisées, ou simplement informés après coup.

Les acteurs impliqués dans la prise de décision forment un organe décisionnel chargé d'établir les normes et politiques pour la gestion des données. Dans les grandes organisations, ces organes peuvent être hiérarchisés, avec différents groupes de travail se concentrant sur des problèmes spécifiques ou des décisions particulières liées aux données. Ces groupes jouent un rôle central en traduisant les besoins stratégiques de l'entreprise en politiques concrètes et en s'assurant que ces politiques sont correctement mises en œuvre à tous les niveaux (Data Governance Institute, s. d.).

Data stewards

Les data stewards, ou responsables des données, sont les "yeux et oreilles" des programmes de gouvernance des données. Ils sont généralement intégrés dans les fonctions métier, ce qui leur permet d'identifier directement les opportunités et les problèmes liés à l'utilisation des données. Leur rôle est essentiel pour assurer que les données répondent aux besoins des unités opérationnelles, tout en respectant les normes de qualité et les politiques de l'entreprise.

Les data stewards sont principalement responsables de la mise en œuvre des contrôles de processus conçus pour standardiser les données, garantir leur qualité et approuver les droits d'accès. Leur travail va au-delà de leurs activités quotidiennes, contribuant de manière significative à la partie opérationnelle de la gestion des données au sein de l'organisation (Data Governance Institute, s. d.).

Data custodians

Les data custodians, ou gestionnaires des données, sont souvent désignés comme les stewards techniques. Ils sont généralement intégrés dans les équipes technologiques ou de gestion des données, où ils appliquent des contrôles, des configurations, et font des choix de conception pour améliorer la probabilité que les données répondent aux attentes des parties prenantes. Leur rôle est vital pour la gestion technique des données, assurant que l'infrastructure technologique supporte les exigences de la gouvernance des données et contribue à la conformité aux normes établies (Data Governance Institute, s. d.).

1.4.5 Défis liés à la mise en place d'un cadre de gouvernance des données

Surmonter la résistance au changement

La mise en œuvre de la gouvernance des données se heurte fréquemment à des résistances, souvent enracinées dans la culture organisationnelle. Ces résistances sont principalement dues à une résistance au changement culturel, un manque de compréhension des bénéfices de la gouvernance des données, ou encore à la peur de perdre le contrôle et l'autonomie sur les données (Sargiotis, 2024). Aborder ces obstacles est essentiel pour une adoption réussie des pratiques de gouvernance des données.

Pour surmonter la résistance au changement dans une organisation, plusieurs stratégies clés peuvent être mises en œuvre, toutes soulignant l'importance d'une communication et d'une gestion de changement efficaces. BorderlessHR (s. d.) en propose plusieurs :

- **Communication efficace** : Il est essentiel de communiquer clairement la vision et les objectifs du changement pour réduire la résistance. Cela permet aux employés de comprendre pourquoi le changement est nécessaire et comment il alignera les objectifs de l'organisation avec les leurs.
- **Engagement des parties prenantes** : Impliquer activement les parties prenantes dès le début du processus de changement favorise un sentiment d'appropriation et de soutien. Cela peut être fait en sollicitant leurs avis et en intégrant leurs feedbacks dans le processus, ce qui réduit considérablement la résistance.
- **Programmes de formation et d'éducation** : Une formation adéquate permet d'équiper les employés des compétences et des connaissances nécessaires pour s'adapter aux nouvelles procédures ou technologies. Le soutien continu tout au long de cette transition est également important pour assurer la confiance des employés dans leur capacité à gérer le changement.
- **Soutien du leadership exécutif** : Le soutien visible et actif des dirigeants est souvent un facteur déterminant du succès du changement organisationnel. Les leaders qui adoptent le changement et le défendent publiquement incitent les autres membres de l'organisation à faire de même.

1.4.6 Exemple de mise en place d'une gouvernance des données

Je vais partager l'exemple d'un programme de gouvernance des données mis en application au sein du Howard County Public School System, développé dans l'article "Data governance, visualization, and utilization: Case studies from four school Districts in various stages of implementation" (Adams et al., 2016).

Le système scolaire public du comté de Howard (HCPSS), situé dans le Maryland, est un système scolaire englobant 76 écoles et plus de 53000 étudiants. En juillet 2013, le HCSP a décidé la mise en place d'un programme de gouvernance des données afin de promouvoir une gestion et une utilisation globale des données de l'organisation pour les responsables scolaires et le personnel de soutien au sein du district. Étant donné le nombre conséquent de parties prenantes de l'organisation pourvoyant des données, l'objectif était d'assurer une gestion plus harmonieuse des différentes données existant au sein de l'organisation.

Ce programme se divise en 10 composantes :

- **Qualité des données :** Maintenir la qualité des données en établissant des objectifs de qualité, des métriques, et des processus pour garantir et évaluer cette qualité. Un contrôle qualité doit être effectué à plusieurs niveaux avant le lancement de tout système.
- **Intégrité des données :** Assurer la cohérence, la fiabilité et l'exactitude des données. Cela implique la mise en place de vérifications automatiques des données entrantes, la définition de règles pour identifier les données mal formatées ou erronées, et la création de mécanismes pour tracer l'origine des données saisies dans le système.
- **Sécurité des données :** Réduire le risque de fuite des données en définissant et en assignant différents niveaux d'accès basés sur les rôles et responsabilités des individus. Le data steward, en charge des audits, rapporte ses découvertes au comité de gouvernance des données.
- **Confidentialité des données :** Préserver la confidentialité des données en s'assurant que l'entreprise respecte les lois et régulations en place. Chaque processus mis en place ou modifié doit être accompagné d'une preuve de conformité.
- **Rétention des données :** Développer une politique claire qui explique et justifie la durée de conservation de chaque type de données au sein des systèmes.
- **Gestion des opérations de données :** Mettre en place des pratiques quotidiennes pour assurer la disponibilité, la sécurité et l'intégrité des données à tout moment.
- **Gestion des données de référence :** Créer un système permettant de structurer et de gérer les données clés nécessaires à la compréhension du fonctionnement d'une organisation. Ces données servent de base pour interpréter et relier les autres données, assurant ainsi cohérence et précision.
- **Gestion du risque :** Mettre en place des stratégies de gestion du risque pour évaluer les vulnérabilités, qu'elles proviennent de menaces externes ou de divulgations non souhaitées. Cela comprend l'identification des données sensibles et la limitation de leur accès.
- **Standardisation :** Assurer la standardisation des données provenant des différentes écoles du district pour garantir la cohérence et l'uniformité des informations.

- **Transparence :** Communiquer les principes fondamentaux de la gouvernance des données et le calendrier aux membres clés du personnel pour assurer une compréhension commune des objectifs et des processus.

Pour assurer le bon fonctionnement de la gouvernance des données, le HCPSS a assigné plusieurs rôles à différents membres de l'organisation à travers une structure très hiérarchisée. Les data owners sont les personnes chargées d'entrer les données dans le système conformément aux principes établis. Ils sont nommés par les data managers, qui sont responsable de résoudre les problèmes concernant la qualité et l'utilisation des données.
Les data stewards sont eux assignés à des domaines spécifiques de l'organisation, et responsables de la qualité et l'utilisation des données de ce département en établissant notamment des droits d'accès, la manière dont les données sont collectées et veillent au bon respect des politiques de gouvernance des données.

Enfin, le comité de gouvernance des données est chargé développer et de gérer le programme de gouvernance des données. Il résout les problèmes opérationnels, nomme les data stewards, collabore avec eux pour s'assurer du bon déroulement de leurs responsabilités, implémente les processus d'amélioration de la qualité des données, et aide à résoudre les problèmes liés aux données.

Le HCPSS a suivi 7 étapes pour mettre en place son programme de gouvernance des données :

1. Évaluer objectivement les domaines clés nécessitant des améliorations
Le comité de gouvernance des données analyse les systèmes en place dans les différents départements et développe une méthodologie de gouvernance des données à l'échelle du district, en s'appuyant sur les processus organisationnels et les pratiques de qualité des données déjà existants.

2. Assurer la disponibilité et l'accessibilité des données
Centraliser les données des différents départements pour qu'elles soient considérées de manière holistique, indépendamment de leur origine départementale.

3. Déterminer les rôles, responsabilités et règles
Une fois les informations centralisées, l'organisation définit les rôles, responsabilités et règles pour les différents membres du personnel impliqués dans la gestion des données.

4.Assurer la qualité des données :
- Profilage des données à l'aide de divers outils pour en tirer des informations pertinentes.
- Analyse des rapports du Département de l'Éducation du Maryland (MSDE) pour vérifier l'exactitude des informations concernant le HCPSS.

- Création de tableaux de bord pour surveiller et assurer la qualité des données stockées.
- Gestion des données dans le temps pour identifier les tendances, les zones d'amélioration et les problèmes de qualité.

5. Mettre en place une infrastructure de responsabilité
Le programme de gouvernance des données établit des responsabilités claires pour le personnel en charge de l'intégrité des données, soutenues par des outils technologiques adaptés.

6. Passer à une culture basée sur les données
Transitionner d'une culture « transactionnelle » (basée sur des décisions à court-terme) à une culture où les décisions sont fondées sur les données disponibles.

7. Construire des mécanismes de feedback efficaces :
Développer des mécanismes de feedback pour soutenir l'amélioration continue des processus, y compris la qualité des données, la sécurité et le respect des réglementations.

Grâce cette gouvernance des données, le HCPSS assure une meilleure qualité des données à disposition pour les parties prenantes. Par ailleurs, l'automatisation et l'uniformisation des processus réduit la charge de travail des membres du personnel. Enfin le programme de gouvernance des données permet une meilleure communication entre les différents départements de l'organisation (Adams et al., 2016).

1.5 Apport des concepts théoriques

Comme souligné dans cette section théorique du mémoire, la possession et la gestion d'actifs de données de haute qualité sont devenues essentielles pour les entreprises. K. N. Preethi et al. (2023) insistent sur l'importance de cette gestion, en particulier pour une startup comme Peppl. Dans leur analyse, ils soulignent que les startups, confrontées à des ressources limitées, doivent optimiser leur performance et allouer judicieusement leurs ressources. Dans ce contexte, l'utilisation efficace des actifs de données devient cruciale pour fournir des informations permettant de soutenir une prise de décision réfléchie et cohérente, minimisant ainsi les risques d'erreurs.

C'est dans cette optique de maximisation de la valeur des actifs de données chez Peppl. que je vais développer, dans la suite de ce mémoire, les différents processus mis en place pour améliorer la gestion des données de la startup, ainsi que proposer un cadre théorique de gouvernance des données adapté à ses besoins spécifiques.

Partie 2 : Description du projet et approche méthodologique

Dans le cadre de ma gestion de projet, mon objectif principal a été d'améliorer les processus de gestion des données au sein de Peppl. Cependant, cet objectif est assez vaste et manque de précision.

Cette partie du mémoire vise donc à définir un cadre clair et structuré pour ce projet, afin de délimiter les aspects sur lesquels concentrer mes efforts et d'identifier des objectifs spécifiques.

2.1 Cadre de la gestion de projet

Présentation de l'entreprise

Peppl. est une startup belge fondée à la fin de l'année 2021 par deux entrepreneuses qui, à la suite d'expériences personnelles vécues notamment lors du confinement, ont pris conscience du poids que représentaient le stress et le manque de confiance en soi sur le bien-être personnel. Après avoir constaté que de nombreuses autres personnes étaient confrontées à des difficultés similaires, les deux entrepreneuses ont décidé d'investir du temps, de l'énergie et des moyens pour aider un maximum de gens dans cette bataille, menée en continu avec soi-même.

Initialement, Peppl se positionne sur le marché avec un modèle commercial combinant produits de soins cosmétiques et moments de bien-être fournis au travers de services numériques. Les produits de soin étaient à l'origine couplés à un code QR donnant accès à une application mobile proposant un programme de bien-être composé de courts exercices audios. Ces exercices, basés sur des techniques scientifiquement prouvées et validés par des professionnels de la santé mentale, sont destinés à être écoutés en même temps que l'application de soins personnels.

L'application Peppl constitue la véritable valeur ajoutée de l'entreprise pour ses clients, se positionnant comme un amplificateur des effets des routines de soins personnels sur le bien-être. Par une approche holistique, la marque propose à la fois son application et ses produits cosmétiques, permettant de transformer ces routines en véritables moments de soin de soi. Ainsi, ces instants privilégiés deviennent propices à l'écoute des exercices audios de l'application, facilitant l'ancrage de ces habitudes bénéfiques au quotidien.

Au cours de l'année 2022, l'entreprise a décidé de recentrer son activité uniquement sur l'application pour offrir la meilleure expérience possible aux utilisateurs. Cette transition s'est traduite par l'abandon de la vente de produits de soins cosmétiques en faveur de partenariats rémunérés avec des marques de soins implantées sur le marché. L'idée est de combiner des produits de routine avec des exercices de bien-être mental de manière numérique. Ces partenariats consistent en l'ajout de codes QR personnalisés sur certains produits des marques partenaires, tels que Decleor (L'Oréal Groupe), MakeSenz ou YuiSkin, permettant à leurs consommateurs d'accéder à l'application Peppl.

Peppl cherche à se distinguer de ses concurrents tels que Clementine, Calm ou Headspace en offrant une approche hybride qui ne se limite pas au virtuel. L'intégration de codes QR sur des produits de soins cosmétiques est au cœur de la philosophie de l'entreprise. Cette approche encourage l'utilisation de l'application lors des moments de soin personnel, maximisant ainsi leur impact.

L'application dans sa dernière version propose un parcours de 21 exercices audio à écouter quotidiennement, ainsi que 5 exercices bonus. Chacun de ces exercices dure entre 3 et 5 minutes. Les exercices sont disponibles uniquement en anglais.
Le public cible de l'entreprise est composé de femmes âgées de 20 à 35 ans et évoluant en milieu urbain. Ce groupe démographique est particulièrement connecté à son téléphone, soucieux de son bien-être et à la recherche de solutions pratiques pour gérer le stress et/ou améliorer la confiance en soi. Ces femmes mènent souvent une vie active et disposent de peu de temps, ce qui les rend réceptives à des solutions courtes et efficaces comme celles proposées par l'application Peppl..

En 2024, alors que l'application comptait environ 450 utilisateurs inscrits, les deux co-fondatrices de Peppl. ont pris la décision d'arrêter les activités de l'entreprise. Plutôt que de continuer à investir dans l'augmentation du nombre d'utilisateurs, elles ont constaté que les partenaires étaient intéressés par l'approche de bien-être mental, mais pas par l'offre telle qu'elle était présentée. Les clients demandaient des services personnalisés, ce qui les a conduit à conclure que le modèle commercial actuel n'était pas viable. Cependant, elles continueront à travailler dans le domaine du bien-être mental, en se concentrant sur l'image de soi et du corps, sous une autre forme de business modèle.

En conséquence, l'application cessera prochainement de fonctionner sur tous les appareils et l'entreprise sera dissoute au cours de l'année 2024.

Spectre de la gestion de projet

Peppl. dispose de trois types de canaux digitaux au sein de son infrastructure, chacun générant des données utilisateurs : son application, son site web, et ses réseaux sociaux (LinkedIn, Facebook, Instagram). Dans le cadre de ma gestion de projet, j'ai choisi de me concentrer spécifiquement sur les processus de gestion des données liés à l'application.

Ce choix s'explique par le fait que l'application est la source de données la plus pertinente pour effectuer des analyses approfondies, permettant de maximiser l'extraction d'informations utiles. De plus, l'application est le seul canal pour lequel Peppl. a accès à des données de première partie, c'est-à-dire des données brutes collectées directement auprès des utilisateurs, contrairement au site web et aux réseaux sociaux pour lesquels l'entreprise n'a accès qu'à des données tierces et agrégées fournies par les plateformes utilisées. Cela rend les données de l'application non seulement plus précieuses, mais aussi plus exploitables pour des analyses détaillées.

Pour les différentes étapes de la gestion des données, j'ai axé mon travail sur la valorisation des données, c'est-à-dire sur l'optimisation de leur utilisation pour en tirer un maximum de valeur ajoutée. Cela implique que je n'ai pas travaillé sur les processus existants liés à la sécurité, l'archivage ou le stockage des données, ces aspects étant déjà bien pris en charge par les systèmes en place. Mon

approche s'est concentrée sur l'amélioration des processus de préparation, d'analyse et de présentation des données pour permettre une utilisation optimale des données.

2.2 Approche méthodologique de la gestion de projet

Maintenant que j'ai posé le cadre de la gestion de projet au sein de Peppl, je peux définir précisément les objectifs fixés pour la gestion de projet. Je vais ensuite développer la méthodologie à appliquer pour atteindre ces objectifs.

2.2.1 Objectif global de la gestion de projet

L'application Peppl. permet aux utilisateurs de créer un compte et de suivre un parcours d'exercices audios quotidiens. Les données collectées via l'application incluent des informations sur le profil des utilisateurs (comme l'âge, le sexe, et la provenance) ainsi que des détails sur leurs interactions avec l'application, tels que les exercices qu'ils suivent et la durée de leur utilisation.

Ces données sont essentielles pour comprendre le profil des utilisateurs, leurs préférences en matière d'exercices, leurs habitudes d'utilisation, et leur engagement global envers l'application. Elles permettent également d'identifier les points de friction ainsi que les aspects les plus appréciés de l'application, en analysant des indicateurs tels que le taux de réalisation des exercices, le temps passé sur chaque session, et les taux de retour des utilisateurs. Ces informations offrent des retours précieux pour mesurer l'engagement et la satisfaction des utilisateurs.

L'objectif principal de cette gestion de projet est de mettre en place des processus robustes permettant d'extraire un maximum d'informations de qualité à partir de ces données. Cela implique non seulement de développer des méthodes pour préparer et analyser les données efficacement, mais aussi d'établir des processus pour présenter et partager ces informations de manière claire et utile aux parties prenantes de l'entreprise.

Cet objectif est important pour l'entreprise, car il contribue directement à sa croissance et à l'amélioration continue de son produit. En optimisant la compréhension des utilisateurs et en améliorant les plateformes de communication, Peppl. peut non seulement fidéliser ses utilisateurs actuels, mais aussi attirer de nouveaux utilisateurs, augmentant ainsi sa visibilité et son impact sur le marché. La mise en œuvre réussie de ces processus devrait permettre à l'entreprise de prendre des décisions plus éclairées et de rester compétitive dans un environnement dynamique.

2.2.2 Sous-objectifs et méthodologie

Tableau 1 : Les sous-objectifs et les méthodes de la gestion de projet

Sous-objectif	Méthode
Prise en main des plateformes de l'infrastructure digitale de Peppl.	La première étape de la gestion de projet a consisté à se **familiariser** avec les différentes plateformes de gestion de données utilisées au sein de Peppl. Pour cela, j'ai commencé par consulter la **documentation interne** de l'entreprise afin d'acquérir une connaissance des outils en place, suivi d'un **entretien** avec mon prédécesseur au poste de responsable des données pour m'expliquer l'utilisation des différentes plateformes. Ensuite, j'ai visionné des **tutoriels** en ligne pour comprendre en détail le fonctionnement de ces plateformes. Enfin, pour maîtriser ces outils, j'ai été chargé de créer un **récapitulatif des performances** de l'entreprise sur ses divers canaux digitaux pour l'année 2023, en utilisant l'outil Tableau.
Amélioration de la qualité des données utilisées	Pour améliorer la qualité des données de l'application, la première étape a été d'améliorer le **script Python** utilisé pour **récupérer** les données stockées dans la base de données Firebase. J'ai ensuite préparé les données récupérées sur l'outil R Studio et mis en place des **analyses** pertinentes des données à l'aide du même outil.
Création des produits de données	À partir des analyses de données effectuées, j'ai créé un **rapport de données** destiné aux parties prenantes internes de l'entreprise. Ce rapport comprend des informations détaillées sur les utilisateurs de l'application ainsi que sur les tendances d'utilisation. Parallèlement, j'ai développé un **tableau de bord** destiné aux partenaires de Peppl, leur permettant d'accéder à des informations pertinentes sur les utilisateurs ayant scanné le code QR de leurs produits. *Une fois le tableau de bord créé, il aurait été nécessaire de **l'intégrer** à l'infrastructure digitale de l'entreprise et de mettre en place un **accès** sécurisé pour les partenaires.*
Documentation des processus	*Une fois les différents processus de gestion des données établis, il aurait été essentiel de **documenter en détail** le fonctionnement et le résultat de chacun de ces processus* *En complément, il aurait également été nécessaire de créer une **documentation globale** répertoriant l'ensemble des processus de gestion des données existants au sein de l'entreprise, tout en expliquant les liens entre ces différents processus.*

Partie 3 : Mise en œuvre du projet

L'objectif principal de cette section du mémoire est de détailler la manière dont ce projet a été réalisé ainsi que les résultats obtenus. Je vais d'abord présenter la situation initiale de la gestion des données au sein de Peppl., puis développer chaque étape de la réalisation de ma gestion de projet.

3.1 Situation initiale

3.1.1 Infrastructure digitale de l'entreprise

Pour obtenir des données en provenance de ses différents canaux digitaux, Peppl. dispose de plusieurs outils au sein de son infrastructure digitale.

Google Firebase

Comme l'explique J. Granados (2020, avril), Google Firebase est une plateforme complète offrant une variété d'outils pour le développement d'applications mobiles et web. Peppl. a intégré Firebase à son application en ajoutant le SDK (Software Development Kit)Firebase, ce qui permet de connecter l'application aux services Firebase. Le SDK Firebase est un ensemble de bibliothèques et d'outils que les développeurs intègrent dans le code de leur application pour interagir avec les services Firebase. Cela facilite le stockage et la synchronisation des données dans le cloud Firebase et permet de suivre les interactions des utilisateurs avec l'application en temps réel.

Google Firebase fournit également à Peppl. un ensemble de graphiques et de statistiques détaillées. Ces graphiques couvrent divers événements liés à l'application, tels que l'ouverture de l'application, le commencement d'un exercice et bien d'autres. Ces visualisations offrent des informations précieuses sur la fréquence et la répartition de ces événements au sein de l'application. Les utilisateurs de Firebase peuvent interagir avec ces graphiques pour comparer différentes périodes ou événements, permettant ainsi une analyse approfondie des données.

En plus de ces outils, la section analytique de Google Firebase fournit à Peppl. des informations essentielles telles que le nombre d'utilisateurs actifs par période, les taux de rétention et l'engagement moyen des utilisateurs. Il est également possible de segmenter les utilisateurs selon divers critères, comme le type d'appareil utilisé, afin de mener des analyses plus ciblées sur les comportements de segments spécifiques de la base d'utilisateurs.

Apple Analytics & Google Play Analytics

Ces deux plateformes permettent à Peppl. d'avoir accès aux données concernant les performances de l'application sur le Google Play Store et l'Apple Store. On peut notamment y retrouver les statistiques de nombres téléchargement par période, la provenance de ces téléchargements ou encore les informations sur le déploiement de mises à jour.

Hovercode

La plateforme Hovercode offre à Peppl. la possibilité de créer et de télécharger des modèles de codes QR. Hovercode fournit également diverses données analytiques sur ces codes QR, permettant de connaître le nombre de scans effectués pour chaque code QR sur une période déterminée. Cela constitue un moyen idéal pour Peppl. de suivre et d'analyser les tendances d'utilisation des différents codes QR.

Google Analytics

Google Analytics est un outil permettant d'analyser les performances et le comportement des utilisateurs sur un site web. Peppl a intégré Google Analytics à son site web en ajoutant un morceau de code JavaScript, appelé balise de suivi, à chaque page de son site web (helloDarwin, 2024)**.**

Une fois Google Analytics implémenté, il collecte une vaste gamme de données sur les visiteurs du site web de Peppl. :

- **Données démographiques** : Google Analytics peut fournir des informations sur l'âge, le sexe et les intérêts des utilisateurs, aidant Peppl à comprendre le profil de son audience.
- **Comportement des utilisateurs** : Les données sur les pages vues, le temps passé sur chaque page, les taux de rebond et les chemins de navigation montrent comment les utilisateurs interagissent avec le site.
- **Acquisition de trafic** : Les sources de trafic (recherche organique, campagnes publicitaires, réseaux sociaux, référents) permettent à Peppl de savoir d'où viennent les visiteurs.
- **Conversions et objectifs** : Peppl peut définir des objectifs (comme les inscriptions à une newsletter ou les achats) et suivre les taux de conversion pour mesurer l'efficacité de ses stratégies marketing.
- **Analyse des appareils** : Les données sur les types d'appareils (mobiles, tablettes, ordinateurs) utilisés pour accéder au site permettent d'optimiser l'expérience utilisateur sur différentes plateformes.

Meta Business Suite

Meta Business Suite est un outil proposé par Meta pour analyser les activités de comptes sur Facebook et Instagram. Grâce à cette plateforme, Peppl. peut accéder à des données d'engagement détaillées, incluant le nombre de likes, de partages et d'interactions sur les publications. Ces informations permettent d'identifier quel type de contenu suscite le plus d'engagement de la part des utilisateurs. De plus, Peppl. peut obtenir des insights sur le profil démographique de ses abonnés, tels que l'âge, le sexe et la localisation, offrant ainsi une meilleure compréhension de son audience.

Linkedin Analytics

LinkedIn Analytics est l'outil fourni par LinkedIn pour permettre aux entreprises d'accéder aux statistiques de leurs pages et contenus. Il offre des informations similaires à celles de Meta Business Suite, incluant des statistiques sur la performance des contenus publiés et des insights sur le profil des utilisateurs. Grâce à LinkedIn Analytics, Peppl. peut suivre l'engagement et l'interaction de ses abonnés, et obtenir des données démographiques et professionnelles sur ses followers.

3.1.2 Utilisation des données

Peppl. exploite les informations obtenues sur ces outils de différentes manières.

Weekly stand-up

Peppl. tient des réunions hebdomadaires avec tous les membres de l'entreprise appelées "Weekly stand-ups". Lors de ces réunions, différentes statistiques de performance de Peppl. sur ses différents canaux sont présentées via l'outil Canva.

Pour l'application, on va présenter notamment :

- Le nombre d'utilisateurs actifs durant la période
- Le nombre d'exercices commencés et terminés
- Le nombre de téléchargements de l'application

Pour les réseaux sociaux, les informations présentées comprennent :

- Le nombre de publications postées sur chaque réseau
- Le nombre moyen d'interactions par publication sur la période
- Le nombre de followers gagnés ou perdus sur la période

Enfin, pour le site web, les informations présentées sont :

- Le nombre de visiteurs uniques du site
- Les pages les plus visitées

On va également ajouter des informations supplémentaires en cas d'activité exceptionnelle, par exemple en donnant les chiffres liés aux mises à jour quand une nouvelle version de l'application est rendue disponible.

Ces informations permettent de comparer les tendances sur différentes périodes afin d'identifier des variations périodiques ou des impacts de campagnes spécifiques. En suivant la croissance ou la diminution des différents indicateurs au fil du temps, on peut mieux comprendre les facteurs influençant les performances. De plus, comparer les chiffres des différents canaux entre eux peut permettre d'observer les éventuels impacts que les activités des différents canaux ont les uns sur les autres.

Création de tableaux de bord

La plupart des outils utilisés par Peppl. proposent des analyses des données collectées, sous formes de visualisations et de statistiques. Ces données peuvent être exportées sous format CSV (Comma-Separated Values). Un fichier CSV est un format de fichier texte simple utilisé pour stocker des données tabulaires, où chaque ligne représente un enregistrement et chaque colonne est séparée par une virgule ou un autre délimiteur (Gavriloff, 2023).

Ces fichiers CSV sont ensuite convertis sous format Excel, puis importés sur la plateforme de visualisation de données Tableau. Peppl. utilise cette plateforme pour créer des tableaux de bord interactifs. Ces tableaux de bord intègrent les données provenant de divers fichiers Excel, permettant une vue d'ensemble cohérente et exhaustive des performances et des tendances.

Tableau offre des fonctionnalités avancées pour analyser et visualiser les données, facilitant l'identification des tendances et d'informations clés. Les tableaux de bord créés sont ensuite partagés avec tous les membres de l'équipe ayant accès au compte Peppl. sur Tableau, assurant ainsi une diffusion efficace et une utilisation optimale des informations disponibles pour soutenir les prises de décisions stratégiques.

Peppl. se sert de Tableau pour créer des visualisations plus complètes et exhaustives des informations de ses canaux digitaux.

Figure 7 : Exemple du parcours des données au sein de Peppl.

Source : Gouverneur S. (2024, 20 juillet). Parcours des données depuis Google Analytics

En plus de ces deux utilisations principales des données, les membres de Peppl. vont également parfois solliciter certaines informations spécifiques au responsable des données dans le cadre de leurs activités.

3.2 Amélioration des processus de gestion de données de l'application

Toutes les données relatives à l'utilisation de l'application sont enregistrées dans la base de données Firebase. À ce stade, Peppl. s'appuie principalement sur les analyses fournies directement par Firebase pour en tirer les informations utiles. Cependant, cette approche présente certaines limites, notamment en ce qui concerne la profondeur et la précision des informations obtenues.

Les analyses fournies par Firebase sont agrégées, ce qui signifie que les données collectées sont compilées pour offrir une vue d'ensemble globale sans exposer les détails spécifiques. Bien que cela puisse fournir une idée générale des tendances, cette méthode d'agrégation dilue les détails importants, rendant plus difficile la détection d'anomalies, d'erreurs ou de sous-tendances subtiles. En conséquence, certaines informations nécessaires pour une compréhension fine des comportements utilisateurs peuvent être occultées.

L'objectif est donc de mettre en place un processus permettant l'extraction directe des données brutes stockées dans Firebase. En accédant à ces données non agrégées, Peppl. pourrait mener des analyses beaucoup plus détaillées et personnalisées, offrant ainsi une meilleure compréhension des comportements utilisateurs et des tendances. Cela permettrait à l'entreprise de prendre des décisions basées sur des informations complètes et précises, renforçant ainsi sa capacité à affiner et à optimiser ses stratégies de développement.

3.2.1 Récupération des données depuis la base des données

Avant de récupérer les données de la base de données, il est nécessaire de comprendre leur format et la manière dont elles sont organisées.

Les données dans la base de données Firebase de Peppl. sont stockées sous forme de documents organisés en collections. Chaque collection regroupe un ensemble d'éléments, caractérisés par des paires clé-valeur spécifiques.

Peppl. gère plusieurs collections dans sa base de données. La première est la collection des exercices, qui contient tous les exercices disponibles sur l'application. Chaque document de cette collection inclut des paires clé-valeur telles que le nom de l'exercice, la collection à laquelle il appartient, et sa durée.

Ensuite, la collection des utilisateurs rassemble toutes les informations relatives aux utilisateurs inscrits sur l'application, comme leur nom, la date de création de leur compte, et leur sexe. Chaque utilisateur a également une sous-collection attribuée qui répertorie les exercices qu'il ou elle a commencés.

Enfin, la collection des événements enregistre toutes les actions possibles sur l'application, comme l'ouverture de l'application, la création d'un compte, ou le lancement d'un exercice. Chaque événement est défini par des paires clé-valeur, notamment le nom de l'événement, la date et l'heure à laquelle il s'est produit, ainsi que l'identifiant de l'utilisateur qui a déclenché l'événement.

Pour extraire ces données, on utilise un script python. Ce script commence par l'importation des bibliothèques Firebase nécessaires pour accéder à la base de données de la plateforme. Les informations d'identification requises sont ensuite chargées pour établir la connexion avec la base de données de Peppl.. Le script initialise ensuite la date d'extraction et crée les listes qui accueilleront les différentes collections de données.

L'étape suivante consiste à parcourir chaque collection pour y stocker, ligne par ligne, toutes les paires clé-valeur de chaque document. Enfin, les doublons éventuels sont supprimés des listes, et les données sont exportées au format CSV.

Il est important de noter que ce script existait déjà chez Peppl.. Mon rôle a été d'y apporter quelques modifications et améliorations pour obtenir des données plus précises et mieux adaptées aux besoins spécifiques de l'analyse.

3.2.2 Préparation des données

L'étape de préparation des données se déroule sur l'outil R Studio, qui facilite l'importation et la modification des fichiers CSV. Le processus consiste à filtrer les données obtenues et à enrichir ces ensembles de données en ajoutant du contexte à certaines variables pour les rendre plus utiles aux analyses. Une fois ce travail de filtrage et d'enrichissement réalisé, les données sont consolidées dans des fichiers Excel. Ces fichiers Excel structurés permettent une analyse plus précise et approfondie, optimisant ainsi l'utilisation des informations collectées pour une plus grande variété d'informations disponibles.

Tableau 2 : Étapes à la préparation des données

Filtrage des données	● Suppression des profils des membres de l'entreprise & des événements associés ● Suppression d'événements sans intérêt pour les analyses futures ● Suppression des doublons des profils utilisateurs ● Suppression des données liées à des mauvaises manipulations des utilisateurs
Enrichissement des données	● Ajout de l'âge des utilisateurs en le calculant à partir de la date de naissance ● Rattachement des événements sans identifiant aux bons identifiants ● Séparation des données concernant le scan des codes QR en fonction de la marque associée ● Ajout de la provenance des utilisateurs à leur profil ainsi que du nombre de codes QR scannés ● Ajout du nombre d'exercices commencés et terminés et du nombre de jours passés sur l'application par utilisateur ● Détermination des jours et heures privilégiés par les utilisateurs pour utiliser l'application
Intégration des données	● Création d'un fichier Excel reprenant les profils d'utilisateurs enrichis

	• Création d'un fichier Excel reprenant les informations détaillées de chaque exercice de l'application • Création d'un fichier Excel reprenant un ensemble d'événements définit comme importants

3.2.3 Analyse et visualisation des données

Une fois les données préparées et intégrées aux divers fichiers Excel, celles-ci sont prêtes pour l'analyse et la présentation. Pour réaliser ces analyses, j'ai utilisé divers outils de statistiques descriptives et de visualisation. Ces outils permettent de transformer les données brutes en insights compréhensibles, en mettant en évidence les tendances, les anomalies et les corrélations essentielles.

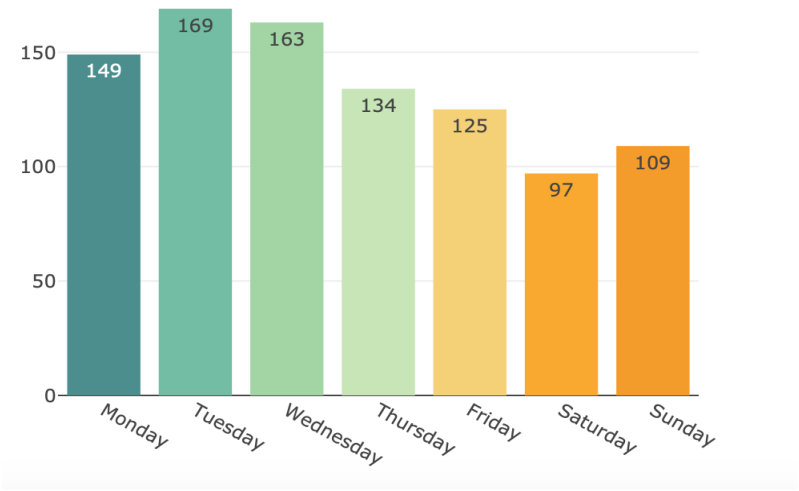
Diagramme en barres

Selon K. Kasmana et F-M. Adipraja (2019), le diagramme en barres est un outil de visualisation des données qui permet de représenter des fréquences ou des quantités relatives à différentes catégories. Cet outil facilite l'estimation et la comparaison des valeurs entre différentes catégories de manière claire et concise.

Un diagramme en barres est constitué de barres rectangulaires dont la longueur est proportionnelle à la valeur qu'elles représentent. Chaque barre correspond à une catégorie distincte, et les barres peuvent être organisées soit verticalement, soit horizontalement.

Les diagrammes en barres sont particulièrement utiles pour visualiser et comparer des données catégorielles. Lorsque les barres ont des hauteurs très variées, cela indique une grande disparité entre les catégories. Inversement, lorsque les valeurs sont proches les unes des autres, les différences peuvent être moins perceptibles, mais le diagramme en barres permet tout de même de les identifier visuellement. Grâce à cette capacité de comparaison visuelle, les diagrammes en barres sont un outil précieux pour l'analyse de données catégorielles, facilitant ainsi une compréhension rapide et intuitive des informations présentées (Kasmana & Adipraja, 2019).

Figure 8 : Nombre d'exercices commencés pour chaque jour de la semaine en 2023



Source : Gouverneur S. (2024, mai). Nombre d'exercices commencés pour chaque jour de la semaine en 2023

Comme le montre la Figure 8, je me suis servi d'un diagramme en barres pour représenter visuellement la répartition des exercices commencés en 2023 sur Peppl. pour chaque jour de la semaine. On peut facilement se rendre compte que l'application est moins utilisée en week-end, tandis que le mardi et le mercredi sont les jours favoris des utilisateurs.

Diagramme circulaire

Un diagramme circulaire permet de représenter des proportions ou des pourcentages relatifs à un ensemble de catégories. Chaque segment du diagramme circulaire représente une catégorie, et la taille de chaque segment est proportionnelle à la valeur qu'il représente, par rapport au total (Yi, s. d.).

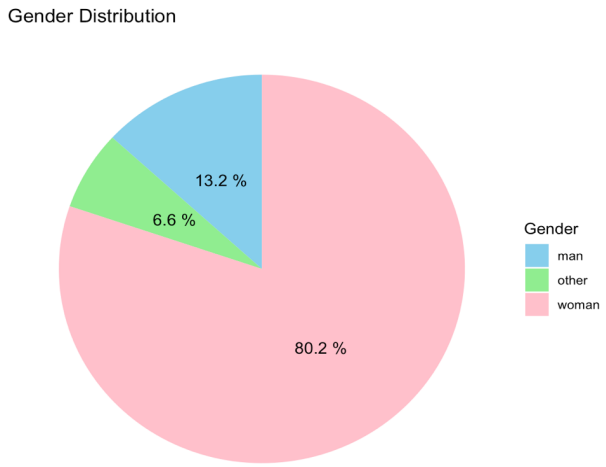
Les diagrammes circulaires sont particulièrement utiles pour visualiser la composition d'un tout et pour comparer les parts relatives de différentes catégories et permettent ainsi de voir rapidement quelles catégories dominent et comment les parts se comparent les unes aux autres.

Selon Yi, M. (s. d.), on va préférer utiliser un diagramme circulaire lorsqu'il est important de montrer la proportion de chaque catégorie par rapport à un tout, comme dans le cas de la répartition des parts de marché entre différentes entreprises ou la proportion des réponses à un sondage. Cela est

particulièrement efficace lorsqu'il y a un nombre limité de catégories et que l'on veut mettre en évidence les parts dominantes.

En revanche, il est préférable d'utiliser un diagramme en barres pour comparer directement les valeurs absolues entre différentes catégories, comme le nombre de ventes par produit ou le nombre de participants à différents événements. Les diagrammes en barres sont également plus adaptés pour visualiser des données avec un grand nombre de catégories et pour montrer des variations importantes entre ces catégories.

Figure 9 : Proportion du genre des utilisateurs de Peppl.



Source : Gouverneur S. (2024, mars). proportion du genre des utilisateurs de Peppl.

Comme on peut le voir sur la Figure 9, ce diagramme circulaire m'a permis de représenter la proportion des différents genres des utilisateurs de Peppl.. Ainsi, on se rend compte qu'une énorme majorité des utilisateurs sont des femmes.

Moyenne et médiane

La moyenne et la médiane sont deux mesures permettant de déterminer la valeur centrale des données quantitatives analysées. Elles permettent d'estimer la tendance générale des données. La moyenne est calculée en additionnant toutes les valeurs de l'ensemble et en divisant cette somme par le nombre total de valeurs.

La médiane, en revanche, est la valeur qui sépare en deux moitiés égales l'ensemble de données. Contrairement à la moyenne, la médiane n'est pas influencée par la présence de valeurs extrêmes dans l'ensemble des données. Si on constate une forte différence entre ces deux mesures, cela indique la présence de valeurs extrêmes dans l'ensemble (Orman, 2022).

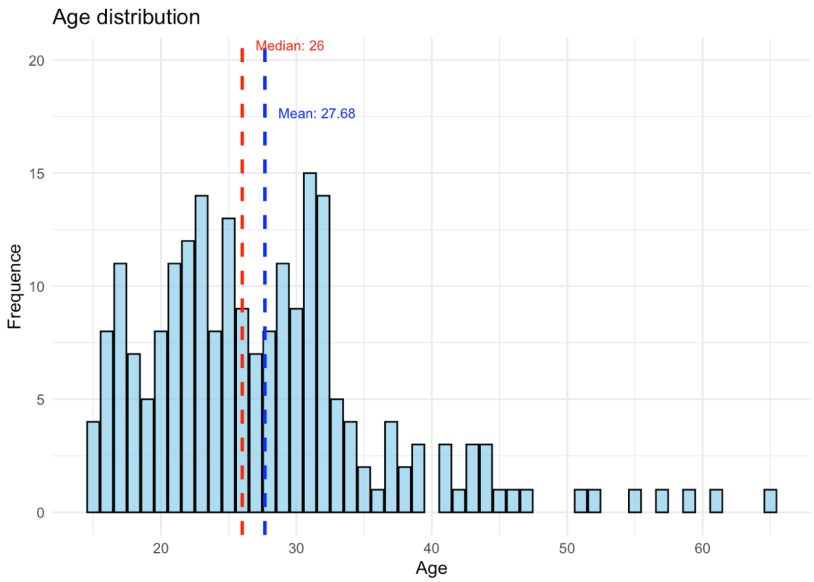
Histogramme

Un histogramme est un outil de visualisation de la répartition de données continues. Il permet d'observer la fréquence des valeurs de cette variable au sein de différentes classes ou intervalles, offrant ainsi une vue d'ensemble de la répartition des données.

Un histogramme est constitué de barres contiguës (sans espace entre elles) dont la hauteur représente la fréquence (ou la densité) des valeurs dans chaque intervalle. L'axe X représente les intervalles de valeurs, et l'axe Y représente la fréquence des observations dans ces intervalles.

En comparant la hauteur des barres, on peut rapidement comprendre où se situent les concentrations de données et détecter des anomalies ou des distributions spécifiques, telles qu'une distribution normale ou bimodale (Coron, 2020).

Figure 10 : Fréquence des âges des utilisateurs de Peppl.



Source : Gouverneur S. (2024, avril). Fréquence des âges des utilisateurs de Peppl.

A moyen de la Figure 10, j'ai représenté à l'aide d'un histogramme la dispersion des âges des utilisateurs de Peppl. On remarque que la majorité des utilisateurs ont entre 20 et 32 ans, avec une

médiane et une moyenne d'âge de respectivement 26 et 27.6 ans. Comme ces valeurs sont très proches, cela indique qu'il y a peu de valeurs extrêmes dans les âges des utilisateurs.

Diagramme de dispersion

Les diagrammes de dispersions sont utilisés pour produire une visualisation qui permet d'identifier visuellement le lien entre deux variables quantitatives. Ce sont des représentations visuelles qui aident à comprendre les relations potentielles entre les variables étudiées.

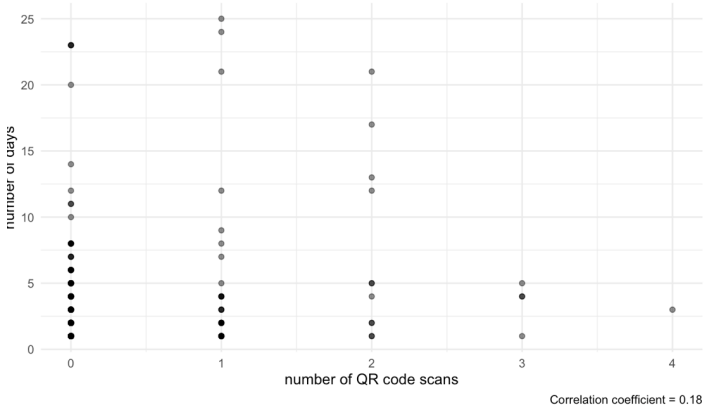
En général on place la variable indépendante (donc qui pourrait influer sur l'autre) en abscisse afin de faciliter l'interprétation.

L'interprétation se fait en observant le positionnement et la dispersion des points. Si on remarque une faible dispersion verticale, alors il est peu probable que les deux variables soient corrélées. Cependant, la corrélation des variables choisies n'indique pas forcément une relation de causalité entre les deux variables, mais potentiellement l'existence d'une variable tierce n'existant pas sur le graphe et néanmoins liée aux deux autres (Coron, 2020).

Coefficient de corrélation

Pour mesurer la relation entre deux variables quantitatives de manière analytique, on va mesurer la corrélation entre les deux. On détermine pour cela un coefficient entre -1 à 1. Si le coefficient tend fortement vers -1, les variables évoluent dans un sens contraire et s'il tend fortement vers 1 elles évoluent dans un même sens. Si le coefficient reste autour de zéro, on ne peut pas dire qu'il y ait de forte corrélation entre les deux variables (Robert, 2023).

Figure 11 : Relation entre le nombre de codes QR scannés et le nombre de jours passés sur l'application



Source : Gouverneur S. (2024, avril). Relations entre l'utilisation de codes QR par les utilisateurs Peppl. et leur fréquentation de l'application

La Figure 11 montre un diagramme de dispersion tentant de montrer la relation entre le nombre de codes QR scannés (de 0 à 4) et le nombre de jours passés sur l'application. Le lien entre ces deux variables semble peu évident visuellement, avec des points de données qui ne semblent pas suivre un ordre précis. Cette impression est confirmée par le coefficient de corrélation des variables qui n'est que de 0,18, pas suffisant pour montrer une relation positive ou négative entre l'utilisation de codes QR par les utilisateurs et leur fréquentation de l'application.

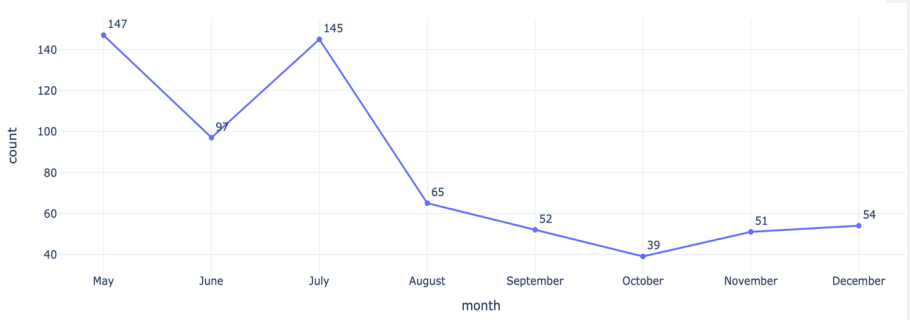
Graphique linéaire

Le graphique linéaire est un outil utilisé pour représenter l'évolution d'une variable quantitative sur une période donnée ou par rapport à une autre variable continue. Les points de données sont connectés par des lignes droites, ce qui permet de visualiser facilement les changements et les motifs dans les données.

Les graphiques linéaires sont particulièrement utiles pour montrer des tendances au fil du temps et aident ainsi à identifier des fluctuations saisonnières ou des anomalies dans les données.

On va utiliser un graphique linéaire lorsqu'il est important de suivre et de visualiser des changements continus et des tendances dans les données (Peters, 2024).

Figure 12 : Nombre d'exercices commencés par mois de mai à décembre 2023



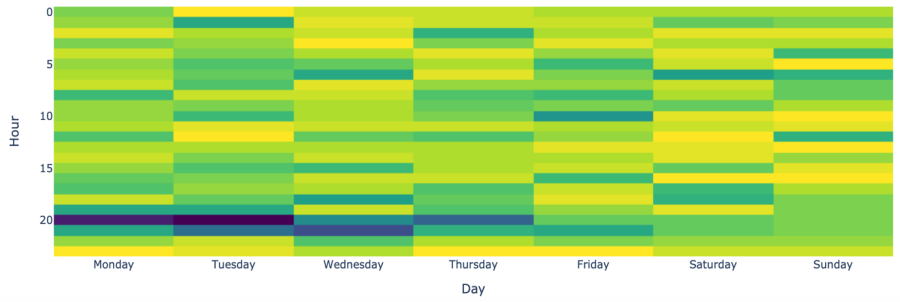
Source : Gouverneur S. (2024, mai). Nombre d'exercices commencés par mois de mai à décembre 2023

La Figure 12 montre un graphique linéaire représentant le nombre d'exercices commencés par mois de mai à décembre 2023. On peut se rendre compte que les utilisateurs ont été très actifs en mai et en juillet, tandis que l'application connaît un creux d'utilisation entre août et décembre de cette année.

Graphique de chaleur

Une carte de chaleur est utilisée pour représenter des données de densité ou d'intensité en utilisant des variations de couleur. La carte de chaleur est particulièrement efficace pour représenter les valeurs d'une matrice où chaque cellule de la matrice a une valeur numérique. Chaque cellule de la carte est colorée en fonction de sa valeur, ce qui permet de voir rapidement les zones de concentration ou les points chauds (Kenton, 2022).

Figure 13 : Nombre d'exercices commencés sur Peppl. pour chaque heure de chaque jour de la semaine en 2023



Source : Gouverneur S. (2024, mai). Nombre d'exercices commencés sur Peppl. pour chaque heure de chaque jour de la semaine en 2023

La Figure 13 est un graphique de chaleur donnant le nombre d'exercices commencés sur Peppl. en 2023 pour chaque heure de chaque jour de la semaine. Plus la case est foncée et plus le nombre d'exercices commencés à cet horaire est important. On se rend ainsi très vite compte que les utilisateurs ont tendance à préférer utiliser l'application en soirée et en début de semaine.

Ces différents outils statistiques, analytiques et de visualisation ont été réalisés dans le cadre de la création d'un rapport des données à l'attention des membres de Peppl., ainsi que d'un tableau de bord destiné aux partenaires commerciaux de l'entreprise.

3.2.4 Création d'un rapport de données

Toujours à l'aide du logiciel R Studio, j'ai créé un rapport de données qui se divise en 3 parties :

- Profils des utilisateurs : leur sexe, leur âge, la plateforme utilisée leur utilisation des codes QR
- Profil des exercices : leur catégorie, leur longueur, et les "moods" auxquels ils sont associés
- Analyses : Cette partie reprend un ensemble d'analyses descriptives, telles que :
    - La distribution des exercices commencés et finis par les utilisateurs,

- La comparaison de l'utilisation de l'application entre les utilisateurs de codes QR et les autres
- Le nombre de lancement et finissement de chaque exercice
- Une analyse des différentes périodes durant lesquelles les utilisateurs effectuent les exercices et s'ils développent une certaine routine

Ce rapport de données fournit une vue d'ensemble générale sur les utilisateurs de l'application et leurs habitudes d'utilisation. Après avoir consulté ce rapport, les membres de l'équipe Peppl. ont pu mieux comprendre le profil des utilisateurs de l'application ainsi que les comportements récurrents.

Ainsi, la présentation de ce rapport a permis certaines observations intéressantes :

- La majorité des utilisateurs de l'application sont des femmes, âgées entre 20 et 33 ans
- La plupart n'utilisent pas de codes QR, mais ceux qui le font ont tendance à plus se servir de l'application
- Environ la moitié des exercices commencés sont terminés par l'utilisateur, et plus d'un quart du total des exercices commencés sont le premier exercice du parcours hebdomadaire
- Une autre observation intéressante que ce rapport montre est que la grande majorité des utilisateurs ayant complété un certain nombre d'exercices ont tendance à se conformer à une routine d'utilisation de l'application

Ces observations permettent de mieux cerner certaines tendances émergentes de l'application, offrant aux décideurs de Peppl. des informations précieux pour optimiser l'application.

Le rapport a été intégré à l'infrastructure digitale de l'entreprise et mis à disposition de tous les membres de Peppl. pour plusieurs raisons. Premièrement, il garantit un accès constant, permettant à chacun de fournir des retours sur les améliorations potentielles ou les ajouts nécessaires. Deuxièmement, ce rapport sert de point de référence temporel, reflétant l'état de l'application pour la période choisie et devant être comparé à des versions futures pour suivre l'évolution de l'activité.

Le rapport se compose principalement d'analyses descriptives et causales, en raison du manque de données suffisantes pour réaliser des analyses prédictives ou de type "what-if". Par exemple, l'application de techniques telles que le clustering, l'analyse de variance ou la régression linéaire multiple n'a pas produit de résultats significatifs en raison de l'insuffisance de données.

Cependant, les fondations établies dans ce rapport pourraient faciliter la mise en place de ces analyses plus avancées en cas d'augmentation substantielle du volume de données. Cela aurait pu permettre à Peppl. de mieux se positionner, en exploitant pleinement des techniques analytiques plus sophistiquées à l'avenir, permettant une compréhension encore plus approfondie des données.

3.2.5 Création d'un tableau de bord

Sur le logiciel PyCharm, j'ai pu créer un tableau de bord interactif affichant diverses données et visualisations représentant les utilisateurs des marques partenaires.

Peppl. collabore avec plusieurs marques de soins cosmétiques qui intègrent des codes QR sur certains de leurs produits permettant d'accéder à l'application. J'ai donc jugé pertinent de développer un tableau de bord personnalisé pour chaque marque.

Ces tableaux de bord fournissent des informations détaillées et pertinentes, permettant aux partenaires de mieux comprendre le succès du partenariat et les habitudes de leurs clients. Grâce à ces visualisations, les marques peuvent suivre des métriques telles que le nombre de scans de QR codes, le taux de conversion des utilisateurs vers l'application, et les interactions au sein de l'application.

Cette approche offre plusieurs avantages :

- **Personnalisation des données** : Chaque marque a accès à un tableau de bord personnalisé qui reflète ses propres utilisateurs et interactions spécifiques.
- **Suivi de performance** : Les marques peuvent mesurer le succès de leur partenariat avec Peppl. en suivant des indicateurs clés de performance.
- **Compréhension des clients** : Les données détaillées aident à identifier les habitudes et préférences des clients, permettant ainsi d'ajuster les stratégies de marketing et de produits en conséquence.
- **Amélioration des partenariats** : En disposant de ces informations, les marques partenaires peuvent collaborer plus efficacement avec Peppl. pour optimiser les campagnes et les interactions avec les clients.

Pour mettre en place ce tableau, j'ai repris les fichiers Excel créés lors de l'étape de préparation, que j'ai importés dans un script Python. A l'aide de bibliothèques telles que Panda, Dash et Plotly, j'ai pu mettre en place diverses analyses et graphiques dans ce tableau.

Ce tableau est divisé en deux tables distinctes :

- La première table reprend des données de base, telles que le nombre d'utilisateurs inscrits via le code QR, le nombre de fois que ce code QR a été scanné ainsi que le nombre total d'exercices commencés et de minutes écoutées sur l'application. On y retrouve également la proportion des genres des utilisateurs, un diagramme circulaire reprenant la distribution des âges, et deux diagrammes en barres illustrant le nombre d'exercices commencés respectivement par période de la journée et par jour de la semaine.

- La deuxième table contient des informations spécifiques à une année ou à un mois. L'utilisateur peut choisir d'afficher des informations pour une année précise ou pour un mois de cette année. On y retrouve le nombre de scans de codes QR et de comptes créés sur la période, un graphique linéaire montrant le nombre de fois que l'application a été ouverte

chaque mois ou jour de la période, ainsi qu'une carte de chaleur permettant de visualiser les ouvertures de l'application pour chaque heure de chaque jour de la semaine.

Étant donné que Peppl. a décidé de cesser ses activités, il n'a pas été nécessaire d'intégrer ce tableau de bord à l'infrastructure digitale de l'entreprise ni de fournir d'accès aux partenaires. Toutefois, voici la méthodologie qui aurait pu être utilisée.

Pour intégrer le tableau de bord à l'infrastructure digitale de Peppl., la première étape aurait consisté à trouver un hébergeur cloud approprié. Il est essentiel de sélectionner un hébergeur qui permet de stocker le script Python ainsi que les fichiers Excel nécessaires. Des options comme Heroku, AWS ou Google Cloud sont recommandées car elles offrent des fonctionnalités de mise en veille automatique pour minimiser les coûts lorsque l'application n'est pas utilisée.

Ensuite, un système d'authentification aurait été ajouté pour restreindre l'accès au tableau de bord. Cela aurait impliqué l'utilisation d'un fichier JSON contenant les informations d'authentification des différents partenaires, incluant les noms d'utilisateur, mots de passe et identifiants de partenaires. Le code du tableau de bord aurait été configuré pour intégrer cette authentification et filtrer les données en fonction de l'utilisateur connecté. Par exemple, lorsqu'un utilisateur se connecte avec le compte de Yuiskin, il ne voit que les données de Yuiskin. Cette étape aurait assuré que chaque partenaire accède uniquement à ses données spécifiques, garantissant ainsi la confidentialité et la sécurité.

La troisième étape aurait consisté à partager les identifiants et mots de passe avec les partenaires concernés, accompagnés d'instructions claires sur la manière d'accéder au tableau de bord et aux informations qu'il contient.

Enfin, il aurait fallu mettre à jour régulièrement les feuilles Excel pour que le tableau de bord affiche des informations à jour. Maintenir des données actualisées est essentiel pour garantir que les informations affichées dans le tableau de bord soient pertinentes et précises pour les utilisateurs partenaires.

En suivant ces étapes, le tableau de bord aurait pu être intégré efficacement à l'infrastructure digitale de Peppl., offrant aux partenaires un accès sécurisé et personnalisé à leurs données actualisées.

Pour proposer cette approche d'intégration du tableau de bord à l'infrastructure digitale de Peppl. et de partage avec les partenaires de Peppl., je me suis aidé du modèle de langage d'intelligence artificielle ChatGPT, développé par OpenAI. La conversation complète a eu lieu le 20 juillet 2024.

3.2.6 Documentation des processus

La dernière étape de ma gestion de projet est une documentation détaillée des processus mis en place lors de la gestion de données, afin de pouvoir facilement s'en servir pour les personnes qui vont me succéder.

La documentation de processus est un ensemble d'instructions écrites avec des descriptions détaillées qui expliquent comment un processus est exécuté au sein d'une organisation, qui ensemble fournissent une base de connaissance, qui contient l'ensemble des documentations de processus d'une organisation. Cette documentation fournit un guide étape par étape de comment réaliser les différentes activités permettant d'atteindre un objectif (Mishra, 2023).

Selon Mishra, T. (2023), une bonne documentation de processus nécessite plusieurs éléments :

- **Impliquer les responsables de processus** : Les personnes participant activement au processus doit participer à la documentation afin de garantir une documentation précise.
- **Simplifier** : Éviter les phrases longues et complexes. Utiliser un langage clair et concis, et structurer les informations de manière simple. Incorporer des images et des diagrammes pour clarifier les concepts. Cela améliorera également la lisibilité de la documentation.
- **Organiser logiquement** : Structurer les documents selon un flux logique qui reflète le déroulement réel du processus. Regrouper les étapes similaires et numéroter les sections et les étapes.
- **Utiliser des modèles** : Recourir à des modèles de documentation de processus pour garantir la cohérence. Compléter ces modèles avec les détails spécifiques au processus.
- **Standardiser le format** : Maintenir la même structure, police, ton et terminologie dans tous les documents à l'échelle de l'organisation pour faciliter la compréhension.
- **Montrer les interactions** : Illustrer les interactions entre systèmes, départements et utilisateurs avec des cartes de processus et des organigrammes.
- **Revoir régulièrement** : Mettre en place des rappels pour revoir la documentation périodiquement afin d'assurer son exactitude, les processus évoluant au fil du temps.
- **Stocker de manière centralisée** : Conserver les documents dans un référentiel central accessible à tout moment par les employés. Partager ces documents entre les différents départements.

En appliquant ces préceptes, la documentation des processus va apporter de nombreux bénéfices au fonctionnement de l'entreprise (Mishra, 2023) :

- **Assurer la cohérence** : Une documentation détaillée des processus permet à chaque employé de suivre les mêmes procédures, réduisant ainsi les variations et les erreurs, tout en maintenant une qualité constante.
- **Préserver les connaissances institutionnelles** : En cas de départ d'un employé clé, la documentation garantit que ses connaissances sont conservées, facilitant ainsi la formation des nouveaux employés et assurant la continuité des opérations.

- **Améliorer l’efficacité** : Des processus bien documentés clarifient les tâches à accomplir, permettant aux employés de travailler plus rapidement et efficacement, ce qui augmente la productivité.
- **Faciliter l’audit** : La documentation des processus simplifie les audits de conformité et de qualité, en offrant une référence claire pour vérifier que les procédures sont bien respectées.
- **Accélérer la formation** : La formation des nouveaux employés est plus rapide et efficace grâce à des documents de référence clairs, garantissant une acquisition de compétences plus uniforme.

Lors de ma gestion de projet chez Peppl., j'ai simplement documenter chaque nouveau processus que j'ai mis en place, en décrivant ce qui était fait et comment chaque étape se déroulait. Mon objectif principal était de fournir une compréhension de base pour chaque phase du processus, afin de rendre l'ensemble plus transparent et accessible pour les utilisateurs futurs. Cependant, cette documentation restait assez générale et ne rentrait pas dans les détails les plus précis, ni des motivations générales des processus.

Si Peppl. avait poursuivi ses activités, il aurait été nécessaire d'aller plus loin dans cette démarche en développant une documentation exhaustive. Cette documentation détaillée aurait couvert l'intégralité des processus, expliquant leur apport à la gestion des données et montrant comment ils s'articulent les uns avec les autres. Elle aurait permis de :

- **Assurer la cohérence** : Une vue d'ensemble des processus aurait facilité la coordination des efforts, en s'assurant que tous les éléments du projet travaillaient ensemble de manière synergique pour atteindre les objectifs fixés.
- **Faciliter la transmission des connaissances :** Pour les futurs membres du département data, une documentation complète aurait fourni une base solide pour comprendre rapidement l'ensemble des processus en place, simplifiant ainsi leur prise en main et leur optimisation.
- **Uniformiser et réduire les erreurs** : Elle aurait garanti que les procédures soient suivies de manière uniforme par tous les membres de l'équipe, réduisant ainsi les risques d'erreurs ou de malentendus.
- **Servir de base pour l'amélioration continue :** Une documentation bien structurée aurait également permis une évaluation continue des processus, facilitant ainsi l'identification des points à améliorer et la mise en œuvre des modifications nécessaires pour optimiser les opérations.

Malheureusement, en raison de l'arrêt des activités de Peppl., cette documentation exhaustive n'a pas été mise en place. Toutefois, l'importance d'un tel document reste évidente pour toute organisation souhaitant améliorer et standardiser ses processus, tout en garantissant la continuité et l'efficacité des opérations.

3.3 Conclusion de la gestion de projet

Cette gestion de projet a permis la mise en place de plusieurs processus de gestion des données de l'application, optimisant ainsi leur utilisation par rapport à la situation initiale. J'ai pu mettre en pratique les différentes étapes d'un cycle de vie des données en les adaptant aux spécificités des données de l'application Peppl..

Ces processus remplissent effectivement l'objectif fixé initialement, qui était de faire en sorte d'obtenir un maximum d'informations de qualité à partir des données, et de permettre aux parties prenantes internes et externes de l'entreprise d'avoir accès aux informations utiles pour eux.

Cependant, l'objectif de ce mémoire dépasse la simple description de ma gestion de projet et la mise en œuvre de processus de gestion des données. Pour maximiser la valeur des données, ces processus doivent être intégrés dans un cadre de gouvernance des données robuste. Un tel cadre ne se contente pas de gérer les données de manière opérationnelle, mais établit également des politiques, des rôles et des responsabilités clairs pour garantir la qualité, la sécurité et l'utilisation stratégique des données au sein de l'entreprise.

Dans la section suivante, je vais explorer la possible mise en place d'un programme de gouvernance des données au sein de Peppl. Je démontrerai comment les processus de gestion des données établis lors de cette gestion de projet peuvent s'intégrer dans un cadre de gouvernance plus large, contribuant ainsi à une meilleure prise de décision et à une optimisation des ressources. Ce programme visera à aligner les données avec les objectifs stratégiques de l'entreprise, à améliorer la transparence entre les départements et à assurer une gestion des données conforme aux normes et réglementations en vigueur.

Partie 4 : Mise en place d'une gouvernance des données

Pour élaborer un cadre de gouvernance des données spécifiquement adapté à Peppl., je m'appuie sur les principes du modèle de gouvernance des données du Data Governance Institute, développé dans la première partie de ce mémoire. Ce modèle, reconnu pour sa rigueur et sa flexibilité, servira de guide pour structurer les processus de gestion des données au sein de l'entreprise.

4.1 Principales sources d'information

Afin de proposer une version adaptée du modèle du Data Governance Institute à Peppl., je me suis basé sur différentes sources.

1. Observation directe

J'ai principalement fondé mon analyse sur des observations directes réalisées lors de mon stage. Il s'agit d'une observation en milieu naturel de type participante.

Ces observations m'ont permis de comprendre en profondeur le fonctionnement de l'organisation, les dynamiques internes, la manière dont les données sont actuellement utilisées, et leur importance stratégique au sein de l'entreprise. Cette immersion m'a offert une perspective unique sur les besoins spécifiques de Peppl. en matière de gouvernance des données.

2. Entretien avec un acteur de terrain

En complément de ces observations, j'ai également mené un entretien approfondi avec un acteur du terrain, maîtrisant les processus de gestion des données mis en place dans son entreprise. Mon interlocuteur était Mr. Benjamin Abdi, cofondateur et chef de produit de Lalilo, une entreprise française spécialisée dans le développement d'une application en ligne destinée à l'apprentissage des compétences en français pour les élèves de 8 à 9 ans. Cet entretien s'est tenu par visioconférence le 26 juillet 2024.

Étant donné que Lalilo, tout comme Peppl., collecte et utilise des données utilisateurs via son application, les processus de gouvernance des données qu'ils ont mis en place présentent des points de repères et de comparaison utiles pour Peppl..

Cette discussion m'a permis de recueillir des exemples concrets de bonnes pratiques à adopter, ainsi que d'identifier les défis spécifiques auxquels une entreprise similaire est confrontée. Ces éléments sont particulièrement pertinents pour la mise en place d'un programme de gouvernance des données au sein de Peppl., car ils offrent des solutions pragmatiques et éprouvées pour améliorer l'efficacité et la sécurité des processus de gestion des données dans un contexte d'entreprise technologique.

Cet entretien a été réalisé avec le consentement éclairé de Benjamin Abdi, c'est-à-dire que l'objectif de l'entretien a été communiqué à l'avance et que l'entretien mené a été enregistré avec son accord.

4.2 Proposition d'un programme de gouvernance des données adapté à Peppl.

Pour mettre en place un programme de gouvernance des données efficace au sein de Peppl., je vais suivre un processus structuré en deux parties.

La première partie consiste à clarifier les objectifs qu'on souhaite atteindre avec ce programme. Il s'agit de définir ce que la gouvernance des données doit accomplir pour Peppl., et la valeur qu'il apporte à l'entreprise. Cette première étape donne une direction claire au programme et permet de s'assurer que tous les efforts sont alignés avec les objectifs stratégiques de l'entreprise. En définissant clairement ces objectifs, on peut mesurer plus facilement le succès du programme et ajuster les efforts si nécessaire.

a supprimé:

La seconde partie a pour but de définir les membres de Peppl. ayant un rôle à jouer dans le cadre d'une gouvernance des données, et ensuite de décrire en détail les étapes importantes pour une mise en œuvre réussie de ce programme de gouvernance des données

4.2.1 Première partie : Définition de la mission et de la valeur ajoutée du programme de gouvernance des données

Pour élaborer un programme de gouvernance des données adapté à Peppl., il est nécessaire de commencer par définir la mission que ce programme doit accomplir et la valeur ajoutée qu'il apporte à l'entreprise. Cette démarche se divise en plusieurs étapes, allant de l'analyse de la situation actuelle à la détermination des attentes et des objectifs futurs.

Situation actuelle de la gestion des données chez Peppl.

Peppl. collecte des données en provenance de son application, ses réseaux sociaux et son site internet. Peppl. collecte des données de première partie directement via son application, ce qui lui permet d'obtenir des informations détaillées sur le comportement et les interactions des utilisateurs. Pour son site web et ses réseaux sociaux, Peppl. ne recueille pas les données directement mais à accès à des données de tierce partie, fournies par des plateformes externes.

Peppl. utilise une infrastructure digitale diversifiée pour gérer et analyser les données provenant de son application mobile, de son site web et de ses réseaux sociaux. Cette infrastructure, détaillée plus en profondeur dans la partie méthodologie de ce mémoire, inclut :

- **Google Firebase**: Gère la base de données des utilisateurs de l'application mobile de Peppl., offrant des services de stockage, synchronisation, et analyse en temps réel.
- **Google Analytics**: Utilisé pour suivre les performances du site web de Peppl., incluant les données démographiques des visiteurs, les sources de trafic, et le comportement des utilisateurs.

- **Meta Business Suite**: Évalue l'engagement sur les réseaux sociaux, notamment sur Facebook et Instagram, en fournissant des insights sur l'interaction des utilisateurs avec le contenu.
- **LinkedIn Analytics**: Analyse l'engagement des utilisateurs sur LinkedIn, offrant des statistiques sur la performance des contenus publiés et le profil des abonnés.
- **Apple Analytics & Google Play Analytics**: Surveillent les performances de l'application mobile sur les App Stores, fournissant des données sur les téléchargements, la provenance des utilisateurs, et l'impact des mises à jour.
- **Hovercode**: Gère les QR codes utilisés par Peppl., fournissant des données analytiques sur leur utilisation, comme le nombre de scans et les interactions des utilisateurs.

Différents processus sont définis et mis en place quant à l'utilisation des données par l'entreprise. Certaines informations de base sont reprises pour les réunions hebdomadaires des membres de l'entreprise, l'outil Tableau permet de créer des tableaux de bord plus complexes à partir de fichiers Excel créés depuis les plateformes de l'infrastructure digitale. Ma gestion de projet a également servi à mettre en place des processus de préparation et d'analyse des données utilisateurs de l'application, ainsi que la création d'un rapport de données récapitulatif des profils et usages de l'application et d'un tableau de bord offrant des informations destinées aux marques partenaires.

Concernant la protection des données, Peppl. a mis en place une politique de confidentialité complète et conforme au RGPD. L'application et le site web demandent notamment l'autorisation aux utilisateurs avant de collecter et traiter les données utilisateur de l'application ou du site web, consentement qui peut être ensuite retiré à n'importe quel moment en contactant l'entreprise (Peppl., 2024).

De plus, les utilisateurs peuvent demander le droit de récupérer une copie des informations personnelles stockées dans la base de données, informations qui 24 mois après la suppression d'un compte utilisateur de l'application, sont supprimées automatiquement de la base de données.

Étant donné que Peppl. ne collecte pas de données sensibles sur ses utilisateurs, telles que des données médicales ou financières, les exigences à respecter en matière de sécurité des données sont moins strictes que celles à mettre en œuvre si ça avait été le cas. Google Firebase assure l'encryption des données, ce qui les protège contre tout accès non autorisé.

Le management des données chez Peppl. est assuré par la personne en charge du département des données. C'est lui qui est responsable de traiter et d'analyser les flux de données entrants, de fournir les informations nécessaires lors des réunions hebdomadaires et de mettre en place des processus d'amélioration des différents aspects propres à la gestion des données. Cette personne est un étudiant-stagiaire.

Les décisions relatives aux données sont prises conjointement par le responsable du département des données et les cofondatrices, principalement autour des aspects de la gestion des données qu'on juge souhaitable d'améliorer.

Points à améliorer dans la gestion de données actuelle

Il serait bénéfique de renforcer la gestion des données chez Peppl. en instaurant des politiques claires et des règles qui standardisent les processus de gestion des données. Actuellement, la responsabilité de la gestion des données est confiée à un étudiant en stage, ce qui entraîne une rotation fréquente des responsables. Chaque nouvel étudiant apporte ses propres méthodes et outils pour gérer les données, sans nécessairement s'aligner sur les pratiques précédemment établies. Ce manque de continuité et de cohérence dans la gestion des données conduit à des inefficacités, à une perte de temps et à une limitation de leur exploitation stratégique. En établissant des protocoles uniformes, Peppl. pourrait non seulement assurer une continuité dans le travail des étudiants, mais aussi améliorer l'efficacité globale de la gestion des données, permettant ainsi une utilisation plus optimale de celles-ci.

Un autre point à améliorer serait de trouver une utilisation plus pertinente des ensemble de données à disposition. À l'heure actuelle, les données sont principalement utilisées pour fournir des statistiques générales sur les performances des différents canaux ou pour offrir une vue d'ensemble de l'utilisation de l'application. Ces informations, bien qu'utiles pour les cofondatrices dans l'évaluation de l'évolution de l'entreprise et de l'application, pourraient être mieux utilisées et à d'autres moments, par exemple pour soutenir la prise de décisions, la stratégie commerciale ou encore le suivi d'objectifs.

Enfin, les données restent relativement inintéressantes pour les autres membres de l'équipe. Or, quel que soit le domaine d'activité de Peppl., les données peuvent représenter un atout certain en tant que soutient des activités réalisées. Il serait donc judicieux de développer des solutions permettant aux autres membres d'avoir accès à des informations plus spécifiques, adaptées aux exigences de leurs projets, soutenant ainsi de manière plus complète et efficace l'ensemble des départements de Peppl.

Objectifs du programme de gouvernance des données

Sur base de la partie précédente, je propose de mettre en place un programme de gouvernance des données avec les objectifs suivants :

- **Assurer la qualité et la continuité des procédés** : Établir des politiques claires et des processus standardisés pour la gestion des données, assurant une continuité malgré la rotation fréquente des responsables des données. Cela permettrait d'éviter des incohérences et une perte de temps liée aux changements fréquents de méthodologies et d'outils, tout en facilitant la transmission des connaissances entre les différents stagiaires et employés. Ce premier objectif représente la gouvernance des données structurelle, et dont les tâches sont assurées par le data steward

- **Création de produits des données** : A partir de chaque ensemble de données collecté, identifier les situations dans lesquelles ces données pourraient apporter une valeur ajoutée, sous quelle forme et comment précisément mettre cela en place. Cet objectif représente la gouvernance des données opérationnelles, et les tâches sont assurées par le data custodian

4.2.2 Seconde partie : Mise en place d'un programme de gouvernance des données

Participants à la gouvernance des données

Au sein de Peppl., l'organisation des rôles liés à la gouvernance des données doit être alignée sur les ressources disponibles et la structure actuelle de l'entreprise. La gestion des données étant principalement assurée par un stagiaire, tandis que les décisions stratégiques sont prises par les deux cofondatrices, la plupart des responsabilités associées au programme de gouvernance des données seront partagées entre ces acteurs principaux.

Acteurs de données et organes décisionnels

Les cofondatrices de Peppl. jouent le rôle central d'organe décisionnel. Elles sont responsables d'établir les normes et les politiques de gestion des données, en s'appuyant sur les informations et les recommandations fournies par la personne en charge des données. Les décisions stratégiques liées aux données, telles que l'alignement des politiques et la définition des priorités, sont prises à ce niveau. Le stagiaire peut également être consulté pour détailler ses analyses et observations, tandis que les autres membres des départements sont informés des décisions prises et peuvent fournir des retours d'expérience basés sur leur utilisation des données.

Data steward

Dans le contexte de Peppl., la personne en charge de la gestion des données agit en tant que data steward. Ce rôle implique de surveiller et de gérer les données au quotidien, de s'assurer de leur qualité, et de veiller à ce que les processus de gestion des données soient conformes aux politiques définies par les cofondatrices. Le stagiaire est également responsable de l'identification des opportunités et des problèmes liés aux données, et de la mise en place de contrôles pour standardiser les données, garantir leur qualité, et réguler les accès.

Data custodian

Bien que Peppl. ne dispose pas d'une équipe technique dédiée spécifiquement à la gestion des données, la personne responsable de la gestion des données peut également assumer une partie des responsabilités de data custodian, en particulier celles qui concernent les aspects techniques de la gestion des données. Cela inclut l'application des configurations techniques et des choix de conception nécessaires pour garantir que les processus de données mis en place sont fonctionnels et permettent d'obtenir ce qu'on souhaite des données. Lorsque des questions techniques plus complexes surviennent, les cofondatrices peuvent décider de faire appel à un support externe ou de faire évoluer le rôle et le profil du stagiaire pour y inclure plus de compétences techniques.

Conception du programme de la gouvernance des données

La conception du programme de gouvernance des données chez Peppl. devrait reposer sur la définition claire et précise de lignes directrices documentées, qui permettraient de réaliser les différentes procédures nécessaires à l'accomplissement des politiques de données mises en place. Cette étape vise à structurer de manière cohérente l'ensemble des actions à entreprendre pour répondre aux objectifs du programme de gouvernance des données.

Assurer la qualité et la continuité des procédés

Pour garantir la qualité, la continuité et la standardisation des données chez Peppl., il est essentiel de mettre en place une approche intégrée qui englobe plusieurs aspects clés de la gestion des données. Cette approche vise non seulement à assurer une exploitation optimale des données disponibles, mais aussi à préparer l'entreprise à une croissance future où les données joueront un rôle de plus en plus central. Pour atteindre cet objectif, il est nécessaire de couvrir plusieurs aspects.

Politiques claires et documentation exhaustive

En premier lieu, il est nécessaire d'établir des politiques claires et uniformes qui régissent l'ensemble des pratiques de gestion des données au sein de Peppl. Ces politiques doivent inclure des lignes directrices précises pour l'utilisation des outils, le stockage des informations, ainsi que pour l'interprétation et la diffusion des données.

En appui à ces lignes directrices, il est nécessaire de développer une documentation exhaustive et dynamique. Cette documentation couvre l'ensemble des plateformes de collecte et de gestion des données présentes dans l'infrastructure digitale de Peppl., ainsi que les différents ensembles de données disponibles et les processus de gestion et d'utilisation des données mis en place.

Cartographie et nomenclature des données

Pour garantir que les données sont à la fois bien connues et de qualité, une cartographie détaillée des données collectées à travers les différents canaux digitaux de l'entreprise est indispensable. Pour l'application et le site web, par exemple, il est pertinent de mettre en place une table de nomenclature des événements. Inspirée des pratiques de Lalilo, cette table répertorie les actions spécifiques des utilisateurs (telles que l'ouverture de l'application, la connexion, ou le clic sur un bouton spécifique du site web) et les données générées par ces actions. Une telle structure offre une vue d'ensemble des événements capturés, les rendant facilement accessibles et compréhensibles pour une exploitation optimale.

Cette approche peut également être étendue aux données collectées depuis les différents réseaux sociaux de l'entreprise, permettant ainsi de disposer d'un registre complet des ensembles de données pour chaque canal digital.

Cette approche permet une vision claire des différents ensemble de données collectés par Peppl..

Contrôles de qualité réguliers

La qualité des données ne peut être assurée sans une vérification régulière à chaque étape des différents processus. Il est donc nécessaire de mettre en place des contrôles systématiques pour identifier et résoudre rapidement tout problème potentiel.

Lors de ces contrôles, on vérifie à l'aide de la table de nomenclature et le registre des données que tous les actifs de données qu'on recense dans ces documents sont effectivement bien collectés, et qu'ils sont aptes à l'utilisation.

On vérifie également les différents processus de gestion des données, et à l'aide de la documentation en place déterminer le bon fonctionnement de ces derniers, et si les produits de donnés en résultants sont bien conformes à ce qu'on attend.

Le data steward en charge d'assurer ces contrôles doit également résoudre les différents problèmes rencontrés lors de ces contrôles.

Transmission des connaissances et mentorat

Enfin, la dernière phase de cette partie du programme de gouvernance des données est d'avoir une phase de remise/reprise du poste de responsable des données la plus efficace possible.

Le contexte de Peppl. fait en sorte que le responsable des données est un poste occupé par un étudiant, un poste qui change souvent de responsable. Ce changement régulier à ce poste peut entraîner une perte de connaissance de la manière donc Peppl. gère ses processus de données.

Afin de soutenir au mieux cette phase de changement, une solution efficace est de faire chevaucher l'arrivée et le départ des responsables des données. Ainsi, cela permettrait d'avoir une période de mentorat, durant laquelle le nouveau responsable des données apprend le fonctionnement des plateformes existantes, les processus mis en place ainsi que les politiques de gestion de données appliquées auprès de son prédécesseur. Cette période permet également de dissiper tout malentendu pouvant exister grâce à la communication entre les deux personnes, et permet de ne pas se reposer uniquement sur la documentation existante pour assurer la transmission des connaissances.

La mise en place de ces différents éléments permet d'encadrer au mieux les différents processus de gestion des données, et d'ainsi permettre une exploitation optimale de ses différents actifs de données.

Création de produits de données

Afin de donner une utilité plus importante aux donnés au sein de Peppl., on envisage à travers le programme de gouvernance des données différents cas d'utilisation des données au sein de l'entreprise, et que mettre en place pour répondre positivement à ces cas d'utilisation.

Aide à la prise de décision

Les données pourraient servir de soutien à la prise de décision. En effet, les co-fondatrices, en tant qu'organe décisionnel de Peppl., peuvent s'appuyer sur des analyses de données détaillées pour orienter leurs choix. Par exemple, si l'entreprise souhaite développer de nouveaux exercices pour l'application, il est intéressant d'utiliser les données liées aux exercices existants, telles que le nombre de fois où ils ont été commencés, le pourcentage de complétion ou encore les « moods » associés à chaque exercice.

En parallèle des données internes, des processus de collecte de données externes peuvent être mis en place pour enrichir les analyses. Par exemple, rechercher des statistiques ou des documents pertinents pour renforcer les bases de décision.

Ces données passent ensuite par des processus de préparation, d'analyse et de visualisation afin de fournir une vision claire et intéressante des informations intéressantes pour la prise de décision.

Un système de feedback pourrait également être mis en place, permettant à l'organe décisionnel de demander des approfondissements sur certaines parties de la présentation, relançant ainsi un cycle d'analyse pour affiner les informations.

Suivi des opérations via des KPIs

Les données pourraient également jouer un rôle central dans le suivi des opérations quotidiennes de Peppl., en particulier à travers la mise en place de Key Performance Indicators (KPIs). Un KPI est défini comme "une mesure quantifiable de la performance sur une période déterminée, liée à un objectif spécifique. Ces indicateurs fournissent des cibles à atteindre pour les équipes, servent de jalons pour évaluer les progrès, et offrent des insights qui aident les personnes au sein de l'organisation à prendre des décisions plus éclairées " (Qlik, s.d.).

Les KPIs sont des outils essentiels pour quantifier la performance de l'entreprise par rapport à ses objectifs stratégiques. Définis en collaboration avec les différents départements, et sur base des performances passées de l'entreprise, ils doivent être en phase avec les objectifs globaux de Peppl. et suivre la nomenclature SMART (Spécifiques, Mesurables, Atteignables, Pertinents, Temporellement définis).

Une fois validés, il incomberait au responsable des données de configurer les outils de visualisation pour créer un tableau de bord interactif, affichant la progression des KPIs en temps réel. Ce tableau de bord serait intégré à l'infrastructure digitale de l'entreprise, accessible à tous les membres concernés et leur laissant la possibilité de commenter ou d'expliquer directement les raisons d'une progression ou d'une baisse des indicateurs.

Les membres de l'entreprise pourront ensuite consulter régulièrement ce tableau de bord pour suivre la progression des objectifs. Le responsable des données serait chargé de mettre à jour ces informations pour s'assurer de la précision des données.

Soutien des activités des différents membres de Peppl.

Les membres de l'équipe Peppl., sont engagés dans diverses activités allant de la création de contenu pour les réseaux sociaux à l'amélioration de l'interface du site web. Pour optimiser ces efforts, il serait pertinent de collaborer avec chaque membre afin d'identifier les domaines où les données pourraient fournir des insights précieux. Cela permettrait de mettre en place un support basé sur les données pour chaque activité, en fournissant par exemple des informations sur les comportements des utilisateurs, les tendances de consommation, ou l'efficacité des campagnes marketing.

En intégrant les données de manière ciblée dans leurs processus, les membres de l'équipe pourraient ainsi bénéficier d'un appui concret pour améliorer la qualité et l'efficacité de leur travail, rendant les actions de Peppl. plus stratégiques et orientées vers les résultats.

Exploitation commerciale des données

Enfin, les données collectées peuvent être exploitées à des fins commerciales, notamment grâce à l'intégration d'un tableau de bord qui compile des informations détaillées sur les utilisateurs des marques partenaires. Ce tableau de bord offre une transparence accrue, permettant aux partenaires de mieux comprendre l'impact et le succès de leur collaboration avec Peppl..

Cette transparence ne se contente pas de renforcer la confiance des partenaires actuels, elle facilite également la renégociation de futurs partenariats. En montrant clairement les bénéfices mutuels, Peppl. se positionne comme un collaborateur de choix, ce qui augmente la probabilité de prolonger ou d'élargir les accords existants.

Ce tableau de bord représente également un atout important pour attirer de nouveaux partenaires. En montrant de manière précise les informations pertinentes que les marques de soins peuvent obtenir en s'associant avec Peppl., l'entreprise présente une proposition de valeur claire et convaincante. Cela rend la plateforme particulièrement séduisante pour les marques désireuses de mieux comprendre et cibler leurs consommateurs, ouvrant ainsi la porte à de nouvelles opportunités de collaboration.

4.3 Limites et projections futures

Bien que ce modèle de gouvernance des données peut être intéressant à mettre en place pour Peppl., il y a certaines limites au bon fonctionnement de ce dernier.

Limitation dans l'application pratique

Le programme proposé reste largement théorique et pourrait s'avérer difficile à appliquer dans la pratique sans ajustements supplémentaires. En effet, la structure de Peppl., avec un seul stagiaire gérant l'ensemble des données et les cofondatrices prenant les décisions stratégiques, ne dispose pas des ressources humaines et techniques nécessaires pour déployer pleinement ce modèle de gouvernance. La mise en place de processus complexes, comme ceux décrits pour la gestion des KPIs, nécessite un investissement conséquent en temps et en compétences que Peppl. n'aurait vraisemblablement pas été en mesure de fournir à court terme.

Manque de ressources techniques et humaines

Le programme de gouvernance des données requiert des ressources humaines et techniques spécialisées, qui étaient limitées chez Peppl. Étant donné que la gestion des données repose principalement sur un stagiaire, il est peu probable que toutes les responsabilités liées aux rôles de data steward et de data custodian puissent être prises en charge avec la rigueur nécessaire. Cela pourrait entraîner une surcharge de travail pour le stagiaire, compromettant ainsi l'efficacité globale du programme.

Faible volume de données disponibles

En raison du volume limité de données actuellement disponible au sein de Peppl., l'impact du programme de gouvernance des données sera probablement modeste à court terme. La place des données resterait assez mineure au sein de l'entreprise, ne pouvant pas fournir autant d'information qu'on le souhaiterait.

Cependant, si l'on envisage que Peppl. mette réellement en place ce cadre de gouvernance des données, il est intéressant d'imaginer comment il pourrait évoluer dans le temps.

Automatisation des processus

À mesure que Peppl. se développe, il serait pertinent d'introduire des outils d'automatisation pour la gestion des données. Par exemple, l'automatisation des audits de qualité, de la mise à jour des KPIs, et des rapports de données pourrait réduire la charge de travail et améliorer l'efficacité. Des plateformes comme Tableau ou Power BI, combinées à des scripts Python, pourraient être utilisées pour automatiser l'analyse et la visualisation des données.

Recrutement d'un spécialiste des données

En cas d'augmentation du volume des données, Peppl. aurait probablement besoin d'engager un spécialiste des données chargé de s'en occuper. Cela serait également l'occasion de lui donner la

responsabilité de la gouvernance des données, et ainsi avoir quelqu'un de plus qualifié en charge. Cela permettrait de répartir les responsabilités, de renforcer la cohérence des processus, et de mieux gérer l'augmentation du volume de données.

Extension des capacités d'analyse

Avoir des volumes de données plus conséquents pourrait permettre à Peppl. d'explorer des analyses plus avancées, telles que les analyses prédictives ou l'apprentissage automatique (machine learning). Ces outils pourraient fournir des informations encore plus précises et aider à anticiper les comportements des utilisateurs, ce qui pourrait être extrêmement bénéfique pour l'optimisation des stratégies de marketing et de développement de produit.

Renforcement des mesures de sécurité des données

Avec l'augmentation du volume de données, il serait important de renforcer les mesures de sécurité des données. Cela passerait notamment par une gestion plus stricte des accès à la base des données, ainsi que par le développement d'un plan d'action à déployer en cas d'incident liés aux données. Peppl. pourrait également envisager de mettre en œuvre, à l'instar de Lalilo, un système de réplication de la base de données, qui se mettrait à jour chaque nuit, garantissant que les analyses de données sont effectuées sur une copie (réplica) et non sur la base de données principale. Ce système permettrait d'éviter que des erreurs de manipulation affectent les opérations critiques de l'entreprise, tout en facilitant des analyses plus rapides et plus sécurisées grâce à des transformations préliminaires des données lors de la réplication.

Conclusion

À partir des recherches effectuées et des résultats obtenus lors de la gestion de projet, j'ai cherché à répondre à la question centrale de mon mémoire :

"Comment améliorer les processus de gestion des données au sein de la startup Peppl. et les intégrer dans un cadre de gouvernance des données ?"

Traditionnellement, les entreprises qui mettent en place un cadre de gouvernance des données sont celles qui traitent un volume massif d'informations, nécessitant une structure rigoureuse pour organiser ces données de manière efficace et en tirer un usage optimal. Ces entreprises ont besoin d'un cadre solide pour s'assurer que leurs données sont exploitées de manière cohérente, sécurisée, et stratégique.

Cependant, le cas de Peppl. est différent. Bien que l'entreprise collecte un volume de données relativement modeste, ce qui pourrait sembler réduire l'urgence d'une gouvernance stricte, l'établissement d'un cadre de gouvernance des données reste néanmoins pertinent. Le cadre que je propose ne vise pas à gérer une grande quantité de données, mais plutôt à instaurer des lignes directrices claires pour la gestion et la valorisation maximales des ensembles de données qu'elle a à disposition, même si elles sont limitées en quantité. Il s'agit d'améliorer la qualité des données, de garantir leur disponibilité, et de maximiser leur valeur en les intégrant de manière stratégique dans les différentes activités de l'entreprise.

Les processus que j'ai mis en place lors de la gestion de projet s'inscrivent directement dans cette démarche. Ils apportent des solutions concrètes aux défis liés à la disponibilité, à la qualité, et à l'exploitation des données de l'application. En établissant des standards pour la préparation et l'analyse des données, en améliorant les mécanismes de suivi des utilisateurs, et en créant des produits de données qui apportent une réelle valeur ajoutée à l'entreprise, ces processus soutiennent et font partie intégrante du programme de gouvernance des données.

En mettant l'accent sur la structuration, la qualité, et l'intégration stratégique des données dès maintenant, Peppl. se prépare à évoluer dans un environnement où les données jouent un rôle de plus en plus central. Ce cadre permettrait à l'entreprise de rester agile, de s'adapter rapidement aux nouvelles opportunités, et de renforcer sa compétitivité à long terme.

D'un point de vue plus global, ce mémoire démontre que l'instauration des principes de gouvernance des données au sein d'une petite entreprise n'est pas seulement pertinente, mais aussi stratégique. Il n'est pas nécessaire de gérer des volumes conséquents de données pour vouloir optimiser leur gestion. En réalité, même une entreprise de taille modeste peut grandement bénéficier d'une gouvernance des données bien pensée, en posant les bases pour une croissance future et en maximisant l'efficacité des processus actuels.

Bibliographie

Abdi, B. (2024, 26 juillet). Chef de produit de Lalilo [Entretien]. Visioconférence

Adams, R. (2016). Data governance, visualization, and utilization: Case studies from four school Districts in various stages of implementation. *SDP FELLOWSHIP CAPSTONE REPORT 2016.* *https://sdp.cepr.harvard.edu/files/cepr-sdp/files/sdp-fellowship-capstone-data-governance-visualization-utilization.pdf*

Alam, I. (2024, 11 mai). *Structured, Semi Structured and Unstructured Data*. K21 Academy. Consulté le 29 juin 2024, à l'adresse https://k21academy.com/microsoft-azure/dp-900/structured-data-vs-unstructured-data-vs-semi-structured-data/#:~:text=Structured%20data%20is%20stored%20in,databases%20or%20other%20data%20table

Aljuwaiber A. (2022, août). Data Warehousing as Knowledge Pool : A Vital Component of Business Intelligence. *International Journal of Computer Science Engineering and Infirmation Technology*, 12(2/3/4), 21-26 https://www.researchgate.net/publication/363402995_DATA_WAREHOUSING_AS_KNOWLEDGE_POOL_A_VITAL_COMPONENT_OF_BUSINESS_INTELLIGENCE

Anter, S., Sassi, I., Bekkhoucha, A. (2021, 29 juin). *A graph-based big data optimization approach using hidden Markov model and constraint satisfaction problem* [Graphique]. Journal of Big Data. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00485-z

Boehm, J., Lewis, C., Li, K., Wallance, D., Dias, D. (2022, Mars 10). *Cybersecurity trends: Looking over the horizon.* McKinsey & Company. Consulté le 28 juin 2024, à l'adresse https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/cybersecurity/cybersecurity-trends-looking-over-the-horizon

Boehm, J., Lewis, C., Li, K., Wallance, D., Dias, D. (2022, Mars 10). *Cybersecurity trends: Looking over the horizon* [Graphique]. McKinsey & Company. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/cybersecurity/cybersecurity-trends-looking-over-the-horizon

Borderlesshr. (s. d.). *10 change management strategies to reduce resistance.* Borderlesshr. Consulté le 3 août 2024, à l'adresse https://borderlesshr.com/blog/change-management-strategies/

Bruno, B., Horn, J. (2022). First party data thriving in the age of privacy regulation. *Applied Marketing Analytics,* 7(2), 211-220. Teradata. https://www.teradata.com/getattachment/eda03b97-9128-4779-90bd-4e969d082d42/first-party-data-thriving-in-the-age-of-privacy-regulation-sp001088.pdf?lang=en-us

Burkhanov K., Jakopin, N., Mohr, G., Cafforio, E., Peres, G. & Weber, M. (2023, Mai). *High data intensity will drive data consumption growth.* Arthur D Little. https://www.adlittle.com/en/insights/report/evolution-data-growth-europe

Chaudhuri S., Dayal U., Narasayya. V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88-98. ResearchGate. https://www.researchgate.net/publication/220422668_An_Overview_of_Business_Intelligence_Technology

China Academy of Information and Communication Technology (2019). *Big Data White Paper.* http://www.caict.ac.cn/english/research/whitepapers/202003/P020200327550643303469.pdf

Coron, C. (2020). Outil 1. Approche quantitative ou qualitative ?. *In La boîte à outils de l'analyse des données en entreprise*. (p. 4-8). Cairn.info. Dunod; Cairn.info. https://www.cairn.info/la- boite-a-outils-de-l-analyse-de-donnees-- 9782100808557-p-20.htm

Coron, C. (2020). Outil 19. Deux variables quantitatives : les nuages de points.
In *La boîte à outils de l'analyse des données en entreprise*. (p. 115-118). Cairn.info. Dunod; Cairn.info. https://www.cairn.info/la- boite-a-outils-de-l-analyse-de-donnees-- 9782100808557-p-20.htm

Coron, C. (2020). Outil 21. Deux variables qualitatives : tableaux et graphiques.
In *La boîte à outils de l'analyse des données en entreprise*. (p. 126-129). Cairn.info. Dunod; Cairn.info. https://www.cairn.info/la- boite-a-outils-de-l-analyse-de-donnees-- 9782100808557-p-20.htm

Cote, C. (2021, Octobre 19). *4 types of data analytics to improve decision-making.* Harvard Business School Online. Consulté le 12 juiller 2024, à l'adresse https://online.hbs.edu/blog/post/types-of-data-analysis

Cote, C. (2021, Décembre 2). *7 data collection methods in business analytics.* Harvard Business School Online. Consulté le 23 février 2024, à l'adresse https://online.hbs.edu/blog/post/data-collection-methods

Cox, M., Ellsworth, D. (1997, Novembre). *Application controlled demand paging for out of core visualization*. [Congrès]. Proceedings of the IEEE 8th conference on Visualization. ResearchGate https://www.researchgate.net/publication/3736976_Application-controlled_demand_paging_for_out-of-core_visualization

Crucial. (2024). *SSD vs. HDD: Know the Difference.* Crucial. Consulté le 19 juillet 2024, à l'adresse https://www.crucial.com/articles/about-ssd/ssd-vs-hdd

Data Governance Institute (s. d.). *Goals and Principles for Data Governance*. The Data Governance Institute. Consulté le 15 juillet 2024, à l'adresse https://datagovernance.com/the-data-governance-basics/goals-and-principles-for-data-governance/

Data Governance Institute (s. d.). *About us*. The Data Governance Institute. Consulté le 23 juillet 2024, à l'adresse https://datagovernance.com/the-data-governance-basics/goals-and-principles-for-data-governance/

Data Governance Institute (s. d.). *The Data Governance Framework and Components.* The Data Governance Institute. Consulté le 18 juillet 2024, à l'adresse https://datagovernance.com/the-data-governance-basics/goals-and-principles-for-data-governance/

Data Governance Institute (s. d.). *The Data Governance Framework and Components* [Graphique]. The Data Governance Institute. https://datagovernance.com/the-data-governance-basics/goals-and-principles-for-data-governance/

Davenport, T., Prusak, L. (1998, Janvier 30) *Working Knowledge: How Organizations Manage What They Know.* Harvard Business Review Press; ResearchGate.

https://www.researchgate.net/publication/229099904_Working_Knowledge_How_Organizations_Manage_What_They_Know

Drexel Lebow's Center for Business Analytics. (2023). *Data integrity Trends and Insights report - Results from a survey of data and analytics professionals*. https://www.lebow.drexel.edu/sites/default/files/2023-06/lebow-precisely-report-2023.pdf

Egidius, P., Abu-Mahfouz, A. M., Hancke, G. P. (2019, Juin). A comparison of data aggregation techniques in software-defined wireless sensor network. *2019 IEEE 28th international symposium on industrial electronics (ISIE),* 1551–1555. IEEE Xplore https://doi.org/10.1109/ISIE.2019.8781537

Foote, K. (2024, 24 avril). *Why Is a data-driven culture important?.* Dataversity. Consulté le 4 août, à l'adresse https://www.dataversity.net/why-is-a-data-driven-culture-important/#:~:text=A%20data%2Ddriven%20culture%20is,on%20previous%20and%20current%20purchases.

Gavriloff, J. (2023, 16 novembre). *Fichier CSV : définition, création et import dans Excel.* https://blog.hubspot.fr/marketing/fichier-csv

Georgiadis, G., Poels, G. (2021). Enterprise architecture management as a solution for addressing general data protection regulation requirements in a big data context: a systematic mapping study. *Information systems and e-business management*, 19, 313-362. Springer Link. https://link.springer.com/article/10.1007/s10257-020-00500-5

Glover, N. (2022, Décembre 19). *What is Data Planning ?.* Quantanite. Consulté le 23 juin, à l'adresse https://www.quantanite.com/blog/what-is-data-planning/

Gouverneur, S. (2024, Janvier - Juin). *Observation du fonctionnement de l'entreprise, de la gestion de données et des dynamiques en place* [Observation directe]. Peppl.. Bruxelles

Granados, J. (2020, Avril 29). *Comment configurer Google Analytics pour Firebase dans votre app ?.* Good Barber. Consulté le 29 juillet 2024, à l'adresse https://fr.goodbarber.com/blog/comment-configurer-google-analytics-pour-firebase-dans-votre-app-a965/

Hannila, H., Silvola, R., Harkonen, J. & Haaspalo, H. (2019, Novembre 25). Data-driven begins with DATA; potential of data assets. *Journal of Computer Information Systems*, 62(1), 29-38. Taylor & Francis. https://doi.org/10.1080/08874417.2019.1683782

Haupt, M. (2016, Mai 2). *"Data is the New Oil" — A Ludicrous Proposition.* Medium. https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294

helloDarwin. (2024, 18 juin). *Comment fonctionne Google Analytics?.* helloDarwin. Consulté le 16 juillet, à l'adresse https://hellodarwin.com/fr/blogue/comment-fonctionne-google-analytics

Hou Z.X. & Xiao, Y. (2019). Big Data for IoT Cloud Computing Convergence. *Web Intellignece*, 17(2), 101-103 https://doi.org/10.3233/WEB-190404

Joshi, S. (2024, Avril 9) *What Is Cloud Storage? 5 Solutions that Deliver High Flexibility.* G2. Consulté le 20 juillet 2024, à l'adresse https://learn.g2.com/cloud-storage

Johnson, E. (2024, juin). Developing scalable data solutions for small and medium enterprises: Challenges and best practices. *International Journal of Management & Entrepreneurship Research,*

6(6), 1910-1935. International Journal of Management & Entrepreneurship Research. https://doi.org/10.51594/ijmer.v6i6.1206

Kasmana, K., Adipraja, F. M. (2019, Novembre). The Benefits of Using Bar Charts in Company Websites. *IOP Conference Series: Materials Science and Engineering,* 662(3). IOP sciences. https://doi.org/10.1088/1757-899X/662/3/032003

Katz, D. (2020, Janvier 23). *What are the 3 V's of big data?* [Image]. Optalitix. https://www.optalitix.com/insights/what-are-the-3-vs-of-big-data

Kenton, W. (2022). *Heatmap: What it Means, How it Works, Example*. Investopedia. Consulté le 20 juillet 2024, à l'adresse https://www.investopedia.com/terms/h/heatmap.asp

Kirvan, P., Wigmore, I., Hashemi-Pour, C. (s.d.) *Definition - What is data lifecycle?.* TechTarget. Consulté le 21 juin 2024, à l'adresse https://www.techtarget.com/whatis/definition/data-life-cycle

Kong, L. & Leil, J. (2020). Fundamentals of big data in radio astronomy. In *Big Data in Astronomy,* (p. 29-58). Elsevier; ScienceDirect. https://www.sciencedirect.com/topics/physics-and-astronomy/big-data#:~:text=In%20the%20same%20year%2C%20the,or%20analysis" %20%5B113%5D.

Laney D. (2001, Février) 3D Data Management: Controlling Data Volume, Velocity, and Variety. *META Group Research*, http://blogs.gartner.com/doug-laney/files/ 2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Lin, L., Liu, T., Zhang, J. (2014). A privacy-aware cloud service selection method toward data life-cycle. *2014 20th IEEE international conference on parallel and distributed systems (ICPADS).* IEEE Xplore. http://ieeexplore.ieee.org/document/7097878/

Lukoianove, T. & Rubin, V. (2014). Veracity Roadmap: Is Big Data Objective, Truthful and Credible? *Advances In Classification Research Online*, 24(1) 4-15. ResearchGate https://doi.org/10.7152/acro.v24i1.14671

Magnisalis I., Peristeras, V., Shah, S. I. H. (2021, Juin). DaLiF: a data lifecycle framework for data-driven governments. *Journal of Big Data*, 8(1) 1-44. ResearchGate https://doi.org/10.1186/s40537-021-00481-3

Magnisalis I., Peristeras, V., Shah, S. I. H. (2021, Juin). DaLiF: a data lifecycle framework for data-driven governments. *Journal of Big Data* [Schéma]. ResearchGate https://www.researchgate.net/publication/352382112/figure/fig3/AS:1034650101227521@1623691215203/DaLiF-Data-Lifecycle-Framework-for-data-driven-governments.png

Mashey R. (1998, Avril). *Big Data ... and the Next Wave of InfraStress.* USENIX. https://static.usenix.org/event/usenix99/invited_talks/mashey.pdf

R.J.T. Morris & B.J. Truskowski (2003, Juillet 1). The evolution of storage systems. *IBM Systems Journal*. https://doi.org/10.1147/sj.422.0205

Mishra, T. (2023, Septembre 8). *What is Process Documentation and Why is it Important?.* Docsie. io Blog. consulté le 28 juillet 2024, à l'adresse https://www.docsie.io/blog/articles/what-is-process-documentation-and-why-is-it-important/

Mistry, S., Prajapati, J., Patel, M., Saxena, S. (2020, Avril). NAS (Network Attached Storage). *International Research Journal of Engineering and Technology* , 7(4), 6571-6575. IRJET. https://www.irjet.net/archives/V7/i4/IRJET-V7I41236.pdf

OpenAI. (2024). ChatGPT. (Version du 20 juillet). [Intégration script Python à une infrastructure digitale]. https://chat.openai.com/chat

Opsmatters. (2024, 12 février). *Data Storage: What Is It and Why Is It Important?*. Opsmatters, consulté le 27 juillet 2024, à l'adresse https://opsmatters.com/posts/data-storage-what-it-and-why-it-important

Orman, E. (2022, septembre 2). *Moyenne par rapport à la médiane*. aide zendesk. https://support.zendesk.com/hc/fr/articles/4408839402906-Moyenne-par-rapport-à-Médiane

Peppl. (2024, 8 mai). *Privacy Policy.* consulté le 20 juillet 2024, à l'adresse https://www.pepplhabits.com/privacy-policy-app

Peters, K. (2024*). Line Graph: Definition, Types, Parts, Uses, and Examples.* Investopedia. Consulté le 20 juillet, à l'adresse https://www.investopedia.com/terms/l/line-graph.asp

Peterson R. (1974, Mars) A Cross Section Study of the Demand for Money: The United States, 1960-62. *The Journal of Finance*, 29(1), 73-88. Google Scholar https://scholar.google.com/scholar?hl=fr&as_sdt=0%2C5&q=Peterson+R.+%281974%2C+Mars%29+A+Cross+Section+Study+of+the+Demand+for+Money%3A+The+United+States%2C+1960-62&btnG=

Petzold, B., Roggendorf, M., Rowshankish, K. & Sporleder, C. (2020, Juin 26). *Designing data governance that delivers value*. McKinsey Digital. Consulté le 16 juillet 2024, à l'adresse https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/designing-data-governance-that-delivers-value

Preethi, K. N. (2023). Enhancing Startup Efficiency: Multivariate DEA for Performance Recognition and Resource Optimization in a Dynamic Business Landscape. *International Journal of Advanced Computer Science and Applications,* 14(8). The Science and Information Organization. https://doi.org/10.14569/IJACSA.2023.0140869

Press, G. (2019, Juillet 17). *A Very Short History Of Big Data.* Forbes. Consulté le 10 juin 2024, à l'adresse https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/

Qlik. (s. d.). *What is a KPI?*. Qlik. Consulté le 25 février 2024, à l'adresse https://www.qlik.com/us/kpi

Rasiff, A. (2023, 11 octobre). T*he pitfalls of a transactional culture: Why your organization needs a shift*. The last eight percent. https://www.ihhp.com/blog/2023/10/11/the-pitfalls-of-a-transactional-culture-why-your-organization-needs-a-shift/#:~:text=At%20its%20core%2C%20a%20transactional,results%20are%20prioritized%20and%20rewarded.

Robert, J. (2023, 1 août). *Coefficient de corrélation : Qu'est-ce que c'est ? À quoi ça sert ?.* DataScientist. https://datascientest.com/coefficient-de-correlation-tout-savoir

Scheuren, F.J., Herzog, T.N., & Winkler, W.E. (2007). What is data quality and why should we care?. In *Data Quality and Record Linkage Techniques*. Springer. https://books.google.be/books?id=iofCetdcJSoC&pg=PA7&redir_esc=y#v=onepage&q&f=false

Santos, M. Y., Sá, J. O., Andrade, C., Lima, F. V., Costa, E., Costa, C., Martinho, B. & Galvão, J. (2017, Décembre). A Big Data system supporting Bosch Braga Industry 4.0 strategy. *International journal of information management,* 37(6), 750-760. ScienceDirect. https://doi.org/10.1016/j.ijinfomgt.2017.07.012

Sargiotis, D. (2024). Data Governance in the Digital Age: Strategies, Challenges, and Best Practices. In Data Governance a guide. (p. 4-49). Springer. ResearchGate. https://www.researchgate.net/publication/377108089

Sargiotis, D. (2024). *Data Governance in the Digital Age: Strategies, Challenges, and Best Practices* [Graphique]. ResearchGate. https://www.researchgate.net/publication/377108089

Shen, Y. (2018). Data Sustainability and Reuse Pathways of Natural Resources and Environmental Scientists. *New Review of Academic Librarianship*, 24(2), 136-156. Semantic Scholar. https://doi.org/10.1080/13614533.2018.1424642

Shi, Y., Shi, H., You, J. & Xu, T. (2023, Décembre). From data to data asset: conceptual evolution and strategic imperatives in the digital economy era. *Asia Pacific Journal of Innovation and Entrepreneurship,* 18(1), 2-20. Emerald insight. https://doi.org/10.1108/APJIE-10-2023-0195

U.S. News (s.d.). *Paterson Public School District*. U.S. News. consulté le 30 juillet 2024, à l'adresse https://www.usnews.com/education/k12/new-jersey/districts/paterson-public-school-district-111799

Villanova University. (2024, Mars 8). *What Is a Business Process? | Definition, Importance and Examples.* Villanova University. Consulté le 25 juillet 2024, à l'adresse https://www.villanovau.com/articles/bpm/what-is-a-business-process/

Vistrada. (2023, Décembre 21). *Data security compliance: standards, regulations, and best practices.* Vistrada. https://vistrada.com/resources/insights/data-security-compliance

Wang, J., Liu, Y., Li, P., Lin, Z., Sindakis, S., Aggarwal, S. (2023, Février).
Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality. Journal of the knowledge economy, 15, 1159-1178. Springer Link. https://link.springer.com/article/10.1007/s13132-022-01096-6

Watson, C. (2019, Juin) From accountability to digital data: the rise and rise of educational governance. *Review of Education*, 7 (2), 390-427. ResearchGate https://www.researchgate.net/publication/327506546_From_accountability_to_digital_data_The_rise_and_rise_of_educational_governance

Yi, M. (s. d.). *How to choose between a bar chart and pie chart.* Atlassian. Consulté le 4 août 2024, à l'adresse https://www.atlassian.com/data/charts/how-to-choose-pie-chart-vs-bar-chart#:~:text=In%20short%2C%20a%20pie%20chart,down%20a%20whole%20into%20components.