Un modèle de protection des données adapté aux nouveaux usages facilités par le *Data Analytics* et le *Cloud*

Présenter un modèle de protection des données apte à répondre aux enjeux de sécurité et de protection de la vie privée.

Ce modèle doit prendre en compte les risques et les nouveaux usages des données associés à la transformation digitale ainsi que l'usage des outils analytiques et du Cloud.

Travaux menés avec

SOMMAIRE

Contributeurs

Xavier BOIDART	Crédit Agricole Assurance
Mohamed BOUTAYBI	Banque de France
Mike CHAPMAN	IBM Security
Bruno DELCROIX	BNP Paribas
Pierre FRESSONNET	Banque de France
Joseph SALAMEH	IBM Security

1. Synthèse managériale

Les entreprises continuent leur transformation digitale vers une organisation où la collecte et l'analyse des données sont devenues un élément crucial de succès. De plus, l'adoption des approches Cloud de type PaaS et SaaS offre de nouveaux services techniques, facilitant ainsi le stockage, le traitement, et le partage des données.

Ces transformations en termes d'usages des données et d'adoption des services cloud introduisent des nouveaux risques associés avec le traitement et le stockage des données.

Il est donc nécessaire d'adapter les mesures de sécurité des données et de protection de la vie privée. Cependant, le choix adéquat de mesures de protection dépend souvent de caractéristiques comme la manière dont sont utilisées les données, les acteurs qui accèdent à celles-ci, ainsi que les services techniques employés.

Tandis que plusieurs modèles de sécurité des données et de protection de la vie privée existent déjà, ils n'abordent pas spécifiquement ces caractéristiques et sont plutôt focalisés sur des mesures de type poste de travail, réseau, identité et authentification.

Par exemple, les modèles de protection de type périmétriques, défense en profondeur, zero-trust, et SASE se concentrent notamment sur l'accès aux ressources en abordant la caractéristique « qui accède ? », mais ne prennent pas en compte d'autres caractéristiques clés qui sont nécessaires pour assurer la protection des données. Par exemple, « quel usage ? », « quel service technique est utilisé ? », et « où se trouve le service à travers lequel les données sont consommées ? ». Les risques et menaces pour les données changent sur la base de toutes ces caractéristiques.

Ainsi, afin de répondre aux enjeux de sécurité et de protection de la vie privée associés avec l'usage des données, nous considérons qu'un nouveau modèle de protection est requis. Nous proposons de structurer ce modèle autour de deux axes :

- Un volet gouvernance décrivant les scénarios métiers d'usage des données ainsi que l'identification des risques associés. Ce volet aura pour objectif de :
  - Définir le processus de traitement d'un nouvel usage métier en termes d'activités, de responsabilité des équipes impliquées, ainsi que d'infrastructure sous-jacente de la gestion des données.
  - Faciliter l'échange et la communication entre les différentes équipes métiers (DPO…), IT, risque et sécurité.
  - Faciliter le respect des réglementations et des politiques internes.
  - Aligner le choix des mesures de protection des données et de la vie privée avec les risques associés aux scénarios d'usage des données sensibles.

- Un volet protection proposant un modèle de contrôles structurés autour de l'infrastructure véhiculant l'accès, le conteneur, et les données elles-mêmes afin de :
  - Fournir un cadre commun de mesures de protection de données, simple et évolutif autour des trois catégories suivantes : **(1)** Réseaux d'accès aux ressources de l'entreprise. **(2)** Fonctions de protection des plateformes et des ressources informatiques (« contenants »). **(3)** Fonctions de protection centrée sur la donnée et sa consommation (« contenu »).
  - Promouvoir la cohérence entre le choix des mesures de protection. Ces mesures de protection ont été regroupées dans un cadre illustré à la page 17 afin de faciliter une cohérence globale entre les différentes équipes.

2. Objectif du document

Les données deviennent de plus en plus déterminantes dans le succès des entreprises.

En effet, ces données facilitent la prise de décision, le développement de produits mieux adaptés aux attentes de leurs clients, et optimisent les processus opérationnels métiers.

Cette transformation de l'usage des données par les métiers implique un focus accru sur l'agrégation et le partage des données ainsi que la gestion de ces données et leur analyse. De plus, l'adoption des approches Cloud de type PaaS et SaaS offre de nouveaux services techniques, facilitant ainsi le stockage, le traitement et le partage des données.

Afin de répondre aux risques introduits par ces transformations et aux réglementations associées, il est nécessaire d'adapter les mesures de sécurité des données et de protection de la vie privée. Cependant, le choix adéquat de mesures de protection dépend souvent de caractéristiques comme la manière dont sont utilisées les données, les acteurs qui accèdent à celles-ci ainsi que les services techniques employés. Il est donc nécessaire de comprendre l'évolution et la variété de ces caractéristiques tout au long du cycle de vie des données ainsi que les obligations réglementaires et légales afin de mieux choisir les mesures de protection des données.

Tandis que plusieurs modèles de sécurité des données et de protection de la vie privée existent déjà, ils n'abordent pas spécifiquement ces caractéristiques et sont plutôt focalisés sur des mesures de type poste de travail, réseau, identité et authentification.

Par conséquent, on tente d'assurer la protection des données à travers une collection de mesures de sécurité de type « *IT infrastructure* » sans comprendre les différents usages métiers de ces données qui influencent les risques et le choix des mesures de protection.

Une nouvelle perspective est donc requise.

Ce document a pour objectif de présenter un nouveau modèle de protection des données apte à répondre aux enjeux de sécurité et de la protection de la vie privée associés à la transformation des entreprises autour des données et du Cloud.

3. Introduction

L'adoption du Cloud et des outils analytiques changent les moyens de stocker, de consommer et de partager les données. Ces évolutions facilitent de nouveaux usages métiers comme la collaboration et l'échange des données ainsi que la combinaison et l'analyse des données.

Ces nouveaux usages sont déterminants pour la réussite de l'entreprise et par conséquent, les métiers attendent un changement de posture de sécurité avec des mesures de protection moins restrictives et aptes à supporter leur transformation.

Cependant ces évolutions associées à la production, au stockage et à la consommation des données créent des risques pour la sécurité et la protection de la vie privée.

Vers une entreprise digitale, centrée sur leurs données

« The goal is to turn data into information, and information into insight. » – Carly Fiorina, former executive, president, and chair of Hewlett-Packard Co (1).

« Data is a key asset for the bank, as all processes rely on data. Value will come from data manipulated, created, stored from and by our services to build services for our clients. » – Bernard Gavgani, BNP Paribas Global CIO (2).

La donnée est devenue un élément crucial dans le succès des entreprises afin d'améliorer leur performance commerciale ainsi que leur efficacité opérationnelle. Les données sur le comportement des clients, leurs préférences et besoins, leur géolocalisation ainsi que les informations sur la marque et sa réputation sont des exemples de données typiquement considérées critiques par des chefs d'entreprises.

La collecte et l'analyse de ces catégories de données permettent, par exemple, aux entreprises de développer de nouvelles propositions de valeur, d'améliorer l'expérience client et d'optimiser les processus métiers. Par exemple :

- Personnaliser les produits et les services offerts – une meilleure compréhension comportementale des clients ainsi que de leurs besoins peut faciliter l'adaptation voire le développement de nouveaux produits ainsi que la personnalisation du prix des services.
- Améliorer l'expérience client du service – l'interaction avec le client peut être adaptée sur la base de leurs préférences et interactions observées pour les différents canaux.
- Améliorer les procédures opérationnelles – l'analyse des interactions entre les clients et les équipes opérationnelles permet d'optimiser voire reconcevoir les processus métiers.

Tandis que l'utilisation des données afin d'améliorer les décisions prises au sein de l'entreprise n'est pas nouvelle, la variété, le volume et la vitesse associés à la collecte et l'analyse des données ont significativement augmenté et posent des enjeux pour la sécurité et la protection de la vie privée.

Le croisement des données issues d'une variété de différentes applications, activités métiers ou écosystèmes peut contribuer à réaliser de meilleurs « *Insights* » à la suite de l'analyse. Cependant le fait de combiner des données au sein d'un même modèle d'analyse nécessite un renforcement des mesures de protection. Par exemple, l'accès aux données brutes doit rester limité en fonction de caractéristiques telles que l'appartenance métiers ou la géolocalisation des acteurs. De plus, la combinaison des données permet potentiellement plus facilement l'identification des personnes et peut donc rendre ces données plus sensibles. Le traitement et l'analyse des données dans un cadre multipartenaires, composé d'entités internes et/ou externes, favorisent la combinaison d'une variété de données. Cependant, les usages associés à cette collaboration introduisent de nouveaux risques à cause, par exemple, du lieu de stockage et du traitement des données ainsi que des effets liés à la combinaison des données.

Plus des données sont collectées, meilleurs sont les résultats d'analyse. Ainsi les mesures de protection doivent être capables de traiter des volumes de données en croissance continue. Ceci peut impliquer, par exemple, la localisation et la tokenisation des données à caractère personnel stockées au sein du Data Lake ainsi que de permettre de répondre aux demandes d'accès (DSAR) dans le cadre du règlement RGPD.

L'adoption des outils analytiques en continu permet aux entreprises d'obtenir des "Insights" en temps-réel d'une activité métier. Le choix des mesures de protection de sécurité et de la vie privée doit en

conséquence être adapté afin d'identifier, classifier et protéger les données sensibles dans un flux continu de données qui sont ingérées dans le Data Hub.

Une évolution des architectures et des usages métier des données

L'ajout de nouvelles fonctionnalités métier implique typiquement l'évolution voire le développement d'une nouvelle application avec un modèle de données et un stockage spécifique.

L'adoption des approches Agile et DevOps réduit les cycles de développement tandis que les architectures micro-services décomposent les applications existantes en services autonomes où chaque service est responsable pour ses propres données implémentées avec des services de type PaaS. En plus de développer leurs propres applications, les entreprises ont tendance aussi à adopter des services SaaS.

La décomposition des applications en micro-services ainsi que l'adoption des services SaaS auront tendance à fragmenter les processus et les données métiers à travers plusieurs environnements cloud souvent sans gouvernance des données homogénéisée. Afin d'adapter les mesures de protection de la sécurité et de la vie privée aux données sensibles traitées et stockées dans ces nouveaux services, il est nécessaire de comprendre les responsabilités des fournisseurs Cloud ainsi que les usages de la donnée facilités par ces services et qui les accèdent.

En complément à ces évolutions, la consommation de données par les différentes applications métiers est en train d'évoluer. Les applications métiers ne sont pas seulement un producteur des données pour les besoins d'analyses, elles intègrent de plus en plus du « *machine learning* » et donc deviennent des consommateurs de données en temps-réel.

Afin de promouvoir une agilité dans le développement des applications intégrant du *machine learning*, les entreprises adoptent des approches de type DataOps. DataOps applique les principes d'Agilité et de DevOps au développement des modèles d'analyses ainsi que la gestion des données opérationnelles utilisées par ces modèles.

D'une manière similaire à DevSecOps, les mesures de protection de sécurité des données et de la vie privée doivent aussi être intégrées dès la phase de conception (« shift left ») et promouvoir l'automatisation. En plus, ces mesures de protection doivent être adaptées aux types de données, leur usage ainsi que les acteurs associés avec les différents environnements de développement du modèle d'analyses.

L'adoption de cloud accélère ces transformations mais introduit de nouveaux risques

L'adoption des services cloud accélère et dépasse un simple usage de IaaS. Ces nouveaux services de type PaaS et SaaS apportent une agilité aux métiers, un accès aux nouvelles technologies ainsi qu'une diversification des moyens de consommer les données.

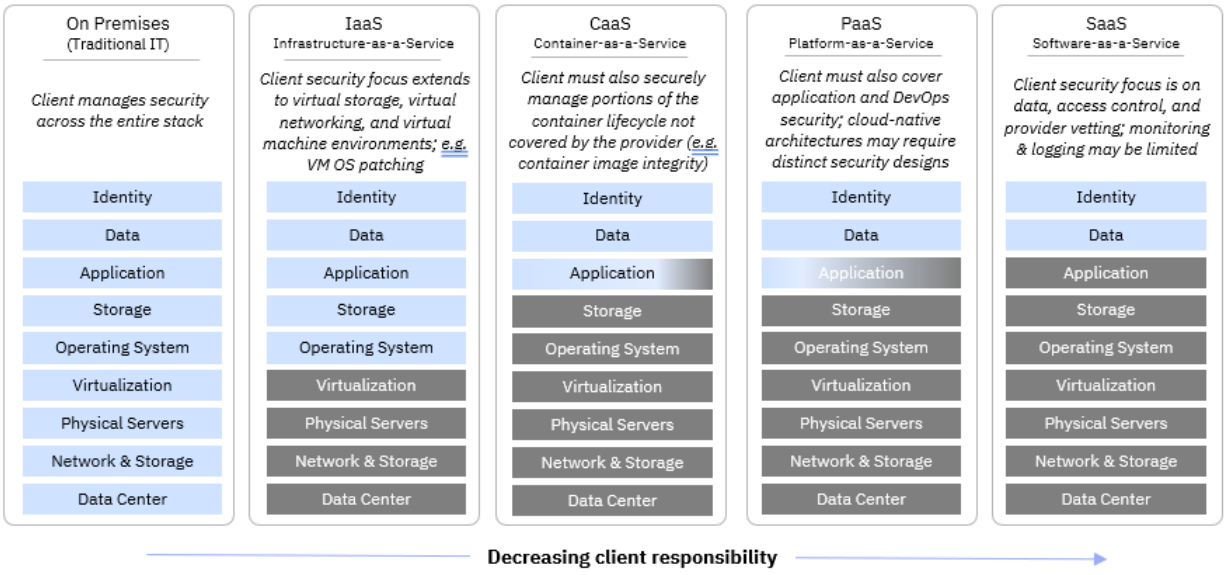
De plus, le Cloud facilite de nouveaux usages, comme la collaboration et l'échange des données. Ces solutions de collaboration et de partage des documents continuent d'évoluer et facilitent de nouvelles façons de travailler pour les utilisateurs (ex : conférence vidéo, messagerie instantanée, édition collaborative). Cependant, il n'est pas évident pour les utilisateurs de savoir s'ils peuvent, ou non, utiliser une solution Cloud en toute sécurité, compte tenu de la sensibilité de certaines données.

L'adoption des solutions SaaS et PaaS, combinée avec un nombre croissant des utilisateurs qui travaillent à distance, a tendance à rendre le périmètre de l'entreprise plus perméable. Les utilisateurs peuvent accéder aux services Cloud sans passer par les Datacenters de l'entreprise et les mesures de

protection associées. En conséquence, les contours de l'entreprises sont moins maîtrisés avec le risque accru, par exemple, de fuite des données.

Ces évolutions impliquent un changement de posture des équipes d'IT et de sécurité. Une approche classique de définition et de déploiement de nouveaux services IT permet d'identifier et d'intégrer des mesures de protection en adéquation avec les risques perçus. Cependant la consommation des services Cloud peut, par leur nature, démarrer très rapidement. En conséquence, cette situation peut entraîner l'usage des services cloud avant même que ces services soient considérés sécurisés par l'ajout des mesures de protection des données complémentaires en fonction des risques. Par exemple, le délai associé avec la mise en place d'une solution de type HYOK « Hold Your Own Key » dans le cadre du chiffrement des données, n'est pas compatible avec la date de début d'usage des services cloud.

Finalement, le Cloud provoque une évolution dans l'écosystème de la production IT, dans les rôles et responsabilités ainsi que dans les lieux d'où l'administration et le support sont effectués. Comme illustré ci-dessous, plus on évolue vers les services de type PaaS et SaaS, moins les équipes clients doivent gérer.



4. Caractéristiques d'un modèle de protection orienté données

Nous considérons qu'un nouveau modèle de sécurité centré sur les données est nécessaire afin de répondre aux enjeux de sécurité des données et la protection de la vie privée.

Difficultés avec les modèles de protection existants (Château Fort, Zero Trust, SASE …)

Les architectures de sécurité réseau sont dans une constante évolution afin de faire face aux transformations dans notre façon de consommer les services IT et les données.

Historiquement, l'architecture de sécurité se basait sur des mesures de protection périmétriques et était principalement orientée autour du réseau et de l'authentification des utilisateurs. Ce type de modèle de sécurité est souvent dénommé « Château Fort ».

Les ressources sont localisées dans les datacenters de l'entreprise et accédées à travers un bastion périmétrique (ex : pare-feu, proxy, VPN, IPS, WAF) qui filtre les flux de communication « nord-sud ». L'accès à la ressource est contrôlé par une combinaison de filtrage, par exemple sur la base de l'IP source de la requête ainsi que les informations d'authentification dont dispose l'utilisateur.

Le modèle représente plus une barrière de passage et d'accès au contenant sur la base d'où la requête vient et qui est l'utilisateur. Cependant avec son focus sur le périmètre de l'organisation, ce modèle ne peut pas répondre aux menaces depuis l'intérieur de l'entreprise.

L'approche de défense en profondeur « Modèle Aéroport » ajoute des mesures de protection en forme d'oignon. Par exemple, les ressources dans les datacenters de l'entreprise sont cloisonnées dans des réseaux virtuels avec un filtrage de communications « est-ouest ». De plus, on analyse les documents importés pour les malwares ainsi que le comportement des utilisateurs ou processus pour la présence d'activités malveillantes.

En synthèse, ce modèle étend le filtrage des flux de communication tout en ajoutant des mesures complémentaires focalisées, par exemple, sur le comportement et la présence de malware. Le modèle reste cependant focalisé principalement sur l'accès au contenant plutôt que sur l'usage des données (le contenu).

Avec l'évolution du cloud et des modèles d'accès distribué, ces modèles et l'architecture associée deviennent de plus en plus inadaptés. Les services de PaaS et SaaS ne sont pas localisés dans les datacenters de l'entreprise et les utilisateurs ont tendance à travailler à distance. En conséquence, les mesures de protection orientées réseau ne sont plus suffisantes puisque l'utilisateur pourra se connecter aux ressources IT à partir du réseau interne, à travers le VPN, ou même directement à travers le réseau internet publique.

De plus l'authentification simple pose un risque pour l'entreprise où les identifiants des utilisateurs pourraient être piratés, ou même partagés entre les utilisateurs.

En conséquence, pour répondre à ces évolutions, un nouveau modèle d'architecture a été mis en avant : le « Zero Trust Network Access » (ZTNA).

Le ZTNA adopte le principe des moindres privilèges et s'appuie sur le concept de « ne jamais faire confiance, toujours vérifier ». En conséquence, pour chaque requête d'accès aux ressources, au sein de l'entreprise ou dans le Cloud, l'utilisateur et son poste de travail possèdent les droits d'accès et attributs appropriés. Ces attributs incluent, par exemple, la géolocalisation, le comportement typique d'accès, ainsi que la version OS et les patches du poste de travail. Le modèle prône également l'usage de l'authentification multi-facteurs afin de renforcer la vérification de l'utilisateur.

Finalement, le modèle d'architecture « Secure Access Service Edge » (SASE), est apparu comme une continuation du modèles ZTNA. L'objectif de SASE est de combiner ces services et technologies pour

créer un réseau cloud sécurisé. SASE prolonge les mesures de protection au-delà des confins de l'entreprise, où ces mesures sont placées au plus près de l'utilisateur au « cloud edge ».

Dans sa présentation du SASE en 2019, Gartner a combiné une multitude de fonctions réseau et sécurité afin de les mettre à disposition des clients comme un service cloud intégré.

SASE combine le SD-WAN ainsi que d'autre services et fonctions réseau incluant :

- ZTNA
- Cloud Access Security Broker (CASB)
- Firewall as a Service
- Secure Web Gateway
- SaaS

Toutefois, à la suite de l'étude de ces modèles d'architecture, nous trouvons qu'ils présentent des limitations vis-à-vis de la problématique de protection des données.

Les architectures ZTNA et SASE se concentrent notamment sur l'accès aux ressources. Ainsi de point de vue des données, ils se concentrent sur la protection du contenant (système de fichiers, base de données …) plutôt que sur le contenu (données sensibles, données règlementées …).

Ainsi la responsabilité du modèle SASE s'avère être limitée à l'accès de l'utilisateur à la ressource. Les mesures de protection sur la manière dont cette ressource et ses données sont ensuite consommées (manipulées) ne sont pas mises en avant. Par exemple une donnée collectée d'une ressource sécurisée pourra être transformée ou partagée vers des destinations externes.

Ces modèles abordent la caractéristique « qui accède ? » mais ne prennent pas en compte d'autres caractéristiques clés qui sont nécessaires pour assurer la protection des données. Par exemple, « quel usage ? », « quel service technique est utilisé ? » et « où se trouve le service à travers lequel les données sont consommées ? ». Les risques et menaces pour les données changent sur la base de toutes ces caractéristiques.

En plus, le traitement des données comme la transformation des données en informations à travers un schéma ou leur combinaison joue un rôle important dans le niveau de sensibilité ainsi que les besoins de protection.

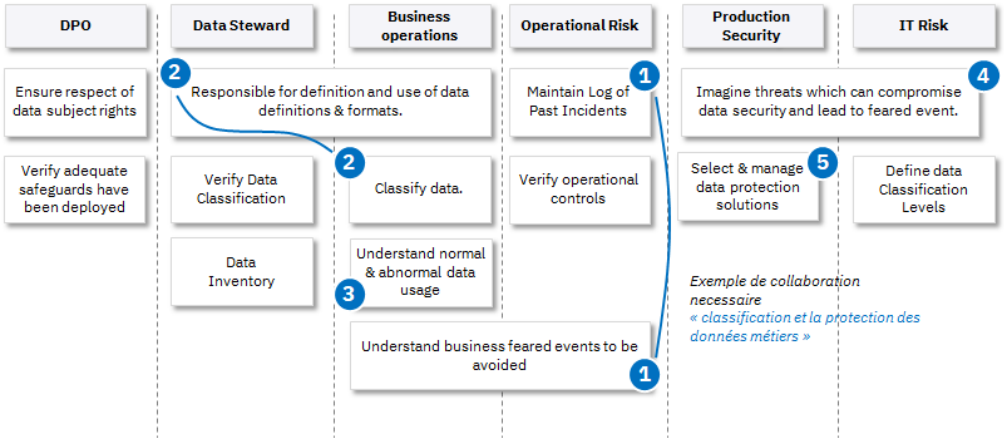
Caractéristiques pour un modèle de protection orienté données

Afin d'éviter les limitations de ces modèles de sécurité existants, nous considérons qu'un modèle de protection des données doit intégrer un nombre de caractéristiques clés afin de répondre aux attentes en termes de sécurité et de la protection de la vie privée.

À un niveau général, le modèle doit fournir un cadre commun, simple et extensible centré sur la protection des données afin de faciliter l'échange et la communication entre les différentes équipes métiers (DPO…), IT, risque et sécurité. Comme illustré ci-dessous, la sécurité des données et la protection de la vie privée nécessitent une collaboration entre différents acteurs à travers l'entreprise afin d'assurer que le niveau de protection est adapté aux risques.

1) Identifier les risques majeurs (évènements redoutés) associés avec les activités métiers en question.
2) Caractériser les données associées avec ces activités métiers (incl. classifier les données, comprendre comment identifier la présence de ces données).
3) Comprendre avec les équipes métiers les usages typiques de ces données et les cas anormaux à éviter.

4) Imaginer comment les évènements redoutés peuvent être déclenchés par différents vecteurs d'attaques.
5) Sélectionner les mesures de protection des données afin de mitiger ces risques.



En plus d'atténuer les risques, le modèle doit également faciliter le respect des réglementations et des politiques internes de l'entreprise en matière de confidentialité, d'intégrité et de disponibilité. Celui-ci implique le besoin d'identifier les données considérées sensibles par les Directives ainsi qu'appliquer des mesures de protection adaptées à ces exigences.

Les données appartiennent aux métiers au sein de l'entreprise. Ainsi, le modèle de protection des données doit être également aligné avec les activités de la gestion des données métiers. Comme décrit par le « Data Management Book of Knowledge » (DMBOK), la sécurité des données et la protection de la vie privée complètent les autres domaines tels que l'architecture des données, les métadonnées et la modélisation des données.

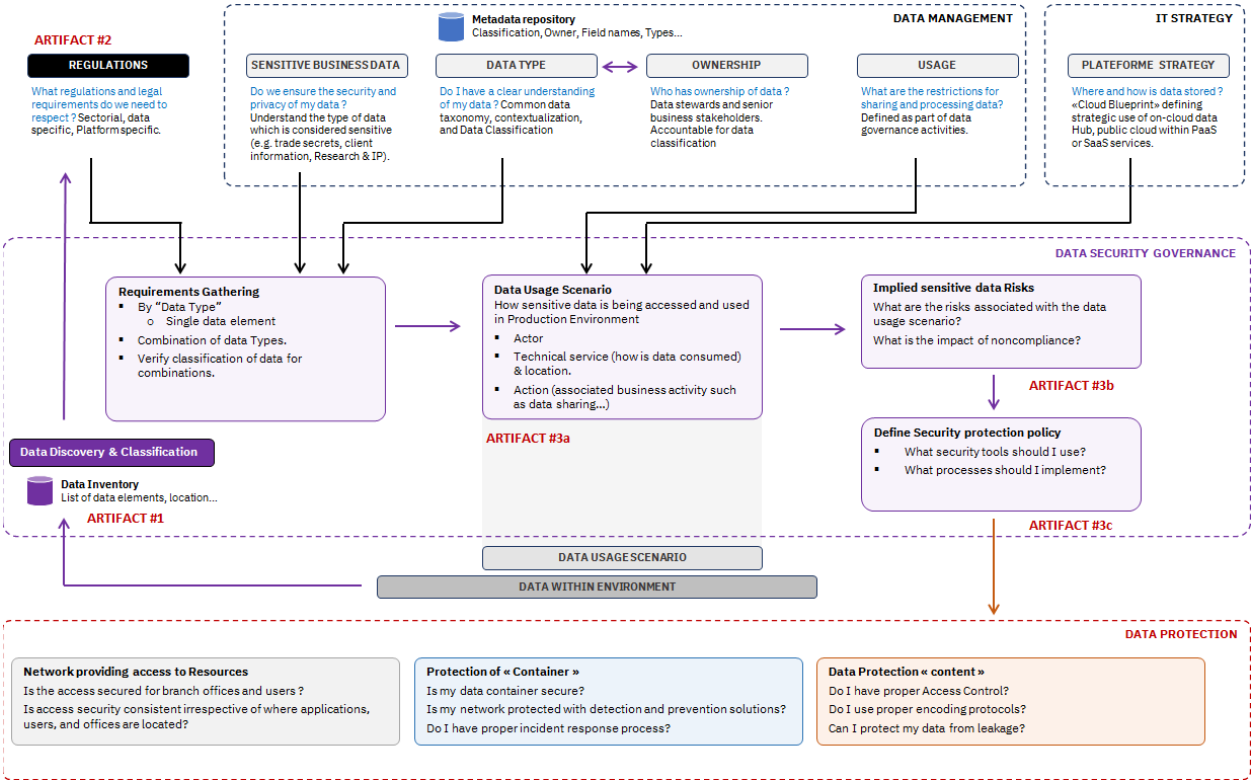
Au-delà de la gestion des données au sein de chaque métier, ce modèle devra prendre en considération la façon spécifique dont les données sont utilisées. Une fois que ces scénarios d'usages (ou de consommation) des données ont été identifiés, ainsi que les acteurs associés, nous pouvons identifier les mesures de protection les plus appropriées.

Le choix des mesures de protection doit également être réalisé en adéquation avec les risques. Ces mesures doivent à la fois permettre un accès approprié aux données, mais aussi protéger la manière dont les données sont consommées.

En plus de fournir un cadre pour l'entreprise pour l'amélioration progressive de la protection des données, nous considérons que le modèle doit également assister les équipes opérationnelles lors de la définition, la mise en place et l'évolution des projets métiers et IT. Ainsi le modèle doit faciliter la cohérence entre le choix des mesures de protection ainsi que l'alignement avec l'architecture de sécurité d'entreprise.

5. Proposition d'un modèle de protection des données

Afin de mieux répondre aux caractéristiques requises pour un modèle de protection des données, nous proposons de combiner une couche de gouvernance avec un cadre de mesures de protection. Ensemble ces deux parties faciliteront l'interaction entre les différentes équipes ainsi que la cohérence du choix des contrôles de sécurité et de la protection de la vie privée.



Un composant Gouvernance.

« Faciliter l'échange et la communication entre les différentes équipes »

La couche de gouvernance a pour objectif de définir les étapes et les responsabilités des différentes équipes impliquées dans la revue des scénarios d'usage des données sensibles, l'analyse des risques et en conséquence la définition des mesures de protection associées.

Le composant gouvernance est fortement dépendant des activités de gestion des données au sein des métiers de l'entreprise. Le catalogue de données jouera le rôle de source de confiance pour ce composant du modèle de sécurité. Ainsi le modèle se basera sur les métadonnées disponibles dans le catalogue pour identifier les données sensibles et en assurer la protection requise.

Dans le cadre de cette gouvernance nous distinguons trois catégories de données sensibles que nous décrivons ci-dessous :

Données Réglementées (Personnelles)

- Avec l'évolution des plateformes de données et l'impact d'une perte ou corruption de ces données sur différents aspects de notre vie quotidienne, les différents gouvernements imposent de nouvelles règlementations pour protéger la vie privée des consommateurs.
- Ainsi au niveau de l'Union Européenne, le Règlement Général sur la Protection des Données (RGPD) promeut un niveau de protection homogène au sein des différents états-membres des données à caractère personnel. Le Règlement assure le respect de la vie privée des personnes concernées à travers l'expression de leurs droits ainsi que des mesures de protection appropriées (ex : purge des données).
- Les obligations exprimées par ces règlementations alimentent les besoins agrégés par l'étape « Requirements Gathering » au sein du composant Gouvernance, tandis que les possibilités de mesures de protection seront reprises dans le catalogue des mesures dans le composant Protection.
- En plus des règlementations, des actes de jurisprudence influencent les règles de protection des données à caractère personnel. Par exemple, « Schrems II », le verdict qui a invalidé le « EU-US Privacy Shield », implique que des mesures supplémentaires sont nécessaires afin d'encadrer les transferts des données à caractère personnel en dehors de l'Union Européenne. https://www.cnil.fr/fr/invalidation-du-privacy-shield-les-consequences-pour-les-organismes-souhaitant-transferer-des

Données Spécifiques aux différents secteurs métiers

- Dans cette catégorie nous distinguons les données spécifiques à l'opération métiers sous-jacente. Par exemple dans le secteur bancaire des mesures devraient être prises pour la protection des données financières.
- En complément des règlementations, des organismes privés positionnent des standards pour gouverner la consommation et la protection de ces données sensibles sectorielles. Par exemple, le PCI / DSS (Payment Card Industry / Data Security Standard) a été établi par les cinq principaux réseaux des cartes du marchés (Visa, MasterCard, American Express, Discover Card et JCB).

Données Spécifiques à l'Entreprise

- Enfin nous identifions les données sensibles spécifiques à l'entreprise et ses plans commerciaux. Nous comptons parmi ces données leur chiffre d'affaires avant publication, les plans de fusion et d'acquisitions, la propriété intellectuelle ainsi que la stratégie interne de l'entreprise.
- La protection de ces données sensibles est vitale pour assurer la croissance de l'entreprise dans un marché compétitif.

Un composant Protection

« Faciliter la cohérence entre le choix des mesures de protection »

Le composant de **Protection** identifie les mesures de protection des données qui sont disponibles au sein de l'entreprise afin de mitiger les risques associés avec le scénario d'usage des données sensibles. Ces mesures couvrent la sécurité des données et la protection de la vie privée.

La combinaison de ces mesures de protection en « design patterns » facilite la compréhension de leur usage tout au long du cycle de vie des données ainsi que la cohérence de choix entre différentes solutions alternatives.

5.1 Composant Gouvernance

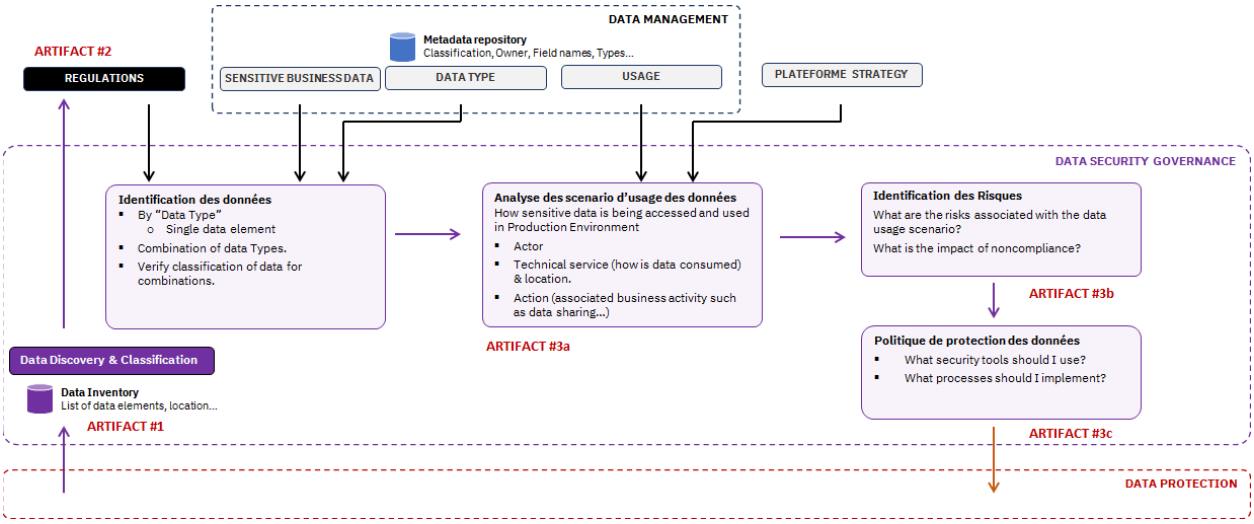
En plus de définir les étapes et les responsabilités associées avec l'analyse des usages des données sensibles et leur protection, le composant Gouvernance a pour objectif de clarifier les dépendances avec la gestion des données au niveau métier.

Dépendances avec la gestion des données métier

« Aligner avec les activités de la gestion des données métier »

Afin d'assurer une gouvernance basée sur la donnée, le modèle de protection des données a besoin de liens étroits avec la gestion des données globale au niveau entreprise et notamment sur le catalogue de données métiers.

En plus, le répertoire central de métadonnées jouera le rôle de source de confiance permettant d'identifier les données ainsi que leur sensibilité et leur propriétaire. Les usages des données peuvent également être progressivement intégrés dans ce répertoire afin de faciliter l'automatisation de la protection des données.



Identification des données
Artifact #1 & #2

La première étape constituant cette gouvernance se concentre sur l'identification des données sensibles consommées par les métiers. Ces données peuvent être considérées comme appartenant à trois principales catégories : des données règlementées qui sont soumises aux lois régionales et locales (ex : RGPD), des données spécifiques au secteur d'activité (ex : PCI/DSS …) et les données métiers de l'entreprise où le niveau de criticité est spécifié par les métiers.

« Faciliter le respect des réglementations et des politiques internes »
Tant les données règlementées et spécifiques au secteur d'activité restent faciles à identifier vue la description disponible dans les textes officiels, tant les données métiers reposent sur la maturité du catalogue de données métiers et l'exhaustivité du répertoire de métadonnées.

Le catalogue fournira un inventaire complet des types de données avec une description permettant de contextualiser cette donnée. Ceci permettra d'aller plus loin dans la compréhension du niveau de criticité de cette donnée en documentant les contextes dans lesquels cette donnée pourra être exploitée comme vecteur de menace pour l'entreprise. Par exemple, du point de vue protection de la vie privée, la civilité ne semble pas être une donnée sensible. Toutefois la relier au prénom ou civilité du conjoint induit l'orientation sexuelle de l'individu considérée sensible.

Les métadonnées mises à disposition peuvent donner plus d'informations sur le propriétaire, le responsable et le format de cette donnée facilitant la découverte des instances de cette donnée dans la plateforme de donnée (« Data Platform »). Ces métadonnées seront configurées dans des moteurs de découvertes permettant de répertorier les instances de ces données sensibles dans l'environnement informatique. Ainsi ces activités de découverte produiront un inventaire de toutes les données sensibles, qui servira de source de confiance pour les composants de protection.

Analyse des scénarios d'usages des données
Artifact #3a

« Scénarios d'usages (ou de consommation) des données »
Une fois les données sensibles bien identifiées, contextualisées et répertoriées, la seconde étape implique l'identification des usages des données conformes aux besoins métiers.

L'objectif de cette étape est de constituer une matrice d'usage reliant les profils d'utilisateurs, les services consommés ainsi que les justifications métiers associées. Ainsi pour chaque donnée sensible, une liste doit être établie montrant pour chaque profil d'utilisateur (Développeur, Analyste, Administrateur…) la justification métiers pour consommer, à travers un service spécifique, une donnée sensible. Par exemple, un analyste devra consommer ces données pour travailler sur des stratégies commerciales, un développeur pour optimiser le code et enrichir les cas d'usage et enfin l'administrateur pour assurer la gestion de cycle de vie et l'hygiène pour le stockage de ces données.

En plus de la justification métier, les canaux de communication pour chacun des profils utilisateurs devront être identifiés en précisant en précisant l'environnement dans lequel la donnée est utilisée (dev, préprod, test, prod …) et la méthode de communication utilisée (accès direct, serveur applicatif …).

Cette matrice permettra une compréhension des différents usages des données afin d'ensuite mieux identifier les risques associés et les besoins de mise en place des mesures de protection. Par exemple, un développeur qui souhaite effectuer des améliorations du code aura besoin d'accéder à un dataset afin de tester son application et la base de données. Toutefois, si ces informations sont anonymisées, il pourra assurer ses activités sans atteinte à la vie privée.

Identification des Risques
Artifact #3b

« Mesures de protection doivent également être implémentées en adéquation avec les risques »

La troisième étape implique la réalisation d'une analyse de risque sur le scénario d'usage des données par les métiers. Cette analyse doit permettra d'identifier les évènements redoutés associés avec les données ainsi que les comportements à risque qui ne sont pas suffisamment contrôlés.

Par exemple, un profil utilisateur ayant les droits d'accès à des données sensibles pourra en abuser d'une façon malveillante ou par étourderie. De plus, ces profils privilégiés restent la cible des pirates informatique menant des techniques d'élévation de privilèges pour avoir accès à des bases de données sensibles.

Parce que ces risques sont spécifiques au métier et associés au cas d'usage, il est recommandé d'aborder chaque scénario d'usage individuellement et ainsi identifier des mesures de protection convenables.

Politique de protection des données
Artifact #3c

Les différentes analyses menées dans les étapes précédentes permettent d'identifier les risques associés avec le scénario d'usage des données sensibles. Ainsi, cette dernière étape permettra d'identifier les différents besoins de protection sur la base de ces risques.

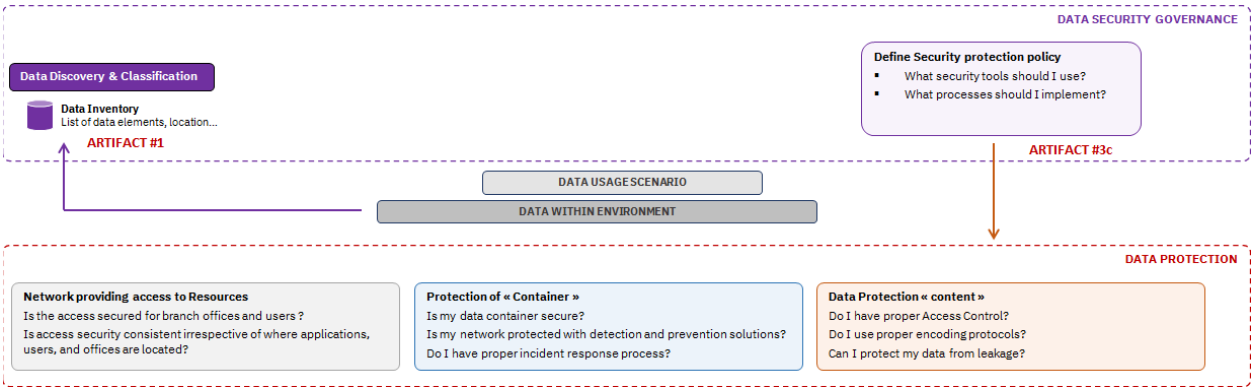
La consolidation de ces besoins de protection permet ensuite de définir une politique de protection appropriée. Cette politique définira les différentes fonctions de sécurité requises afin d'assurer une protection de bout en bout des données sensibles. En plus des fonctions et contrôles de sécurité à appliquer, cette politique inclura aussi les différentes procédures fonctionnelles pour assurer la gestion du cycle de vie de ces données sensibles depuis leur création jusqu'à leur destruction. Cette Politique jouera ensuite le rôle de feuille de route pour le composant « Protection » du modèle de sécurité.

Afin de faciliter la réutilisation des mesures de protection, celles-ci peuvent être regroupées dans des « Patterns de Protection » qui promeuvent la cohérence dans le choix des solutions ainsi que l'industrialisation de la protection des données en termes de sécurité de la vie privée.

5.2 Composant Protection

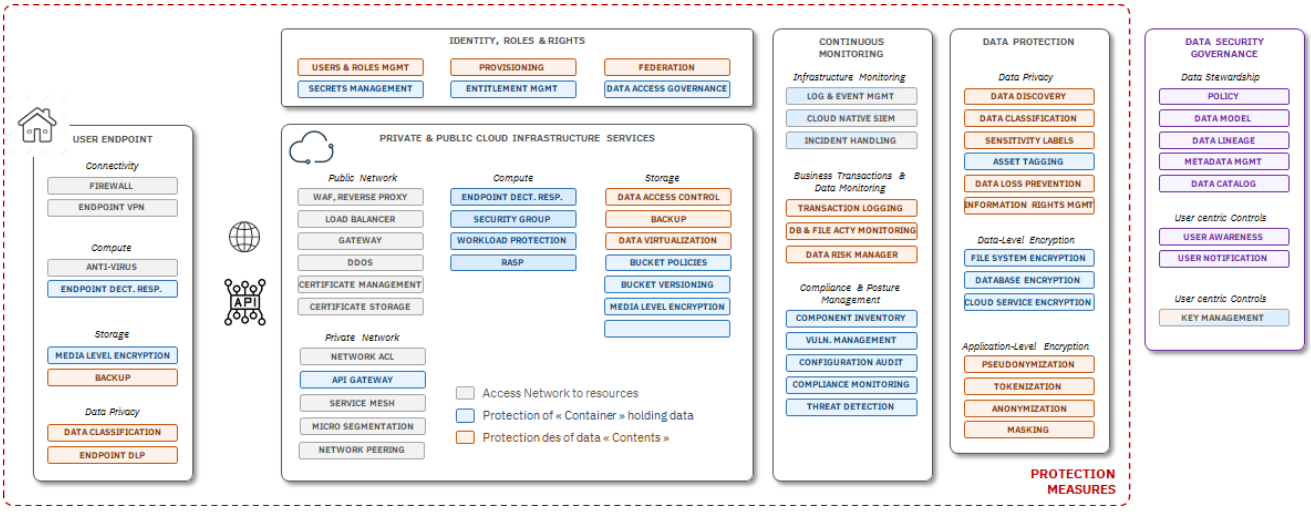
Le composant « Protection » assure la réalisation et la mise en vigueur des mesures de protection de la sécurité des données et de la vie privée. Ce composant se base sur la politique de protection, définie au niveau du composant gouvernance, pour identifier les différentes fonctions de sécurité requises afin d'assurer une protection des données sensibles au sein du cas d'usage.

Afin d'appliquer les règles de protection définies aux données sensibles, il est nécessaire d'identifier leurs localisations sur les différentes plateformes de traitement et de stockage. Ainsi, le composant de protection s'appuie sur l'inventaire des données sensibles qui permet d'identifier la distribution de ces données à travers les ressources informatiques utilisées par l'entreprise.



Les mesures de protection associées avec ce composant ont été séparées en trois catégories. La première catégorie des mesures focalise sur les réseaux d'accès aux ressources de l'entreprise. La deuxième catégorie des fonctions de protection concerne les plateformes et les ressources informatiques où les données sont stockées (« contenants »), alors que la troisième catégorie offre des fonctions de protection centrée sur la donnée et sa consommation (« contenu »).

Ces mesures de protection sont illustrées dans le cadre ci-dessus et décrites dans les paragraphes suivants.



Réseaux d'accès aux ressources « contenants »

Les mesures de protection de la catégorie « réseaux d'accès aux ressources » sont focalisées sur la sécurisation de la connectivité des utilisateurs aux ressources de l'entreprise. Ces ressources peuvent être au sein des datacenters gérés par l'entreprise ou dans des environnements cloud externes de type IaaS, PaaS, ou SaaS.

Ces mesures de protection ont pour objectif principal de faciliter et sécuriser l'accès à ces contenants en menant la connectivité de quelqu'un jusqu'à la porte de la ressource. L'approche SASE (« Secure Access Service Edge ») répond principalement aux besoins associés avec cette catégorie et inclut, par exemple, des solutions de filtrage périmétrique de type pare-feu ou WAF ainsi que des bastions pour les administrateurs.

Protection des « Contenants »

Cette catégorie de mesures est focalisée sur la mise en place de mesures de protection pour les contenants de données. La notion de « contenant » concerne les différents moyens où les données, structurées ou non-structurées, sont stockées. Ceux-ci incluent, par exemple, les dossiers sur le système des fichiers, les buckets au sein du Cloud Object Storage (COS), un Data Lake, ou des bases de données.

Les mesures de protection ont pour objectif de contrôler l'accès à ces contenants ainsi que d'assurer un niveau de sécurité adéquat. Ainsi, différentes mesures de protection, telles que la revue des configurations, l'analyse de vulnérabilités, le chiffrement niveau système des fichiers, « Asset Tagging – classification des assets IT » et le filtrage réseau proche des ressources comme les « Security Groups », doivent être mises en service pour assurer la sécurité de ces contenants avant de commencer l'ingestion des données sensibles.

En plus de la protection de cette infrastructure, qui va « contenir » les données, il faudra assurer une intégration étroite avec les plateformes de sécurité au niveau entreprise notamment avec les solutions de détection et de gestion des incidents pour intégrer la protection des données dans les scénarios de réponses aux incidents.

Protection des données « contenu »

Une fois que l'infrastructure IT est sécurisée, des mesures de protection, plus granulaires et centrées sur les données, peuvent être mises en place afin de renforcer la protection des données sensibles. Ces mesures focalisées sur la sécurité des données et la protection de la vie privée peuvent être regroupées en deux catégories où leur emploi sera gouverné par la politique de sécurité issue de la gouvernance.

1. **Gestion des politiques de protection des données « Policy Decision Point »** a pour objectif d'opérationnaliser la politique de protection des données associées avec le scénario d'usage. Cela concerne la définition et la mise en place des règles de contrôles d'accès, pour les profils utilisateurs et administrateurs (« Users & Roles Mgmt »), selon le contenu accédé ainsi que la détection de comportements contraires à cette politique (« Database & File Activity Monitoring »).

   Ainsi pour un administrateur base de données, nous pouvons bloquer toutes les commandes de types DML (Data Manipulation Language) sur les données sensibles afin de les protéger des

accès non conformes à la politique, tout en permettant à cet administrateur d'accomplir ses tâches administratives habituelles.

Des exemples de ce type de mesures de protection incluent :

- Data Discovery & Data Classification. Par exemple, ces mesures permettent de générer une alerte quand un document sensible est détecté sur un dossier non compatible avec la politique.
- Database & File Activity Monitoring. File Activity Monitoring (FAM) et Database Activity Monitoring (DAM) effectuent la surveillance des activités des administrateurs et des utilisateurs, permettant ainsi de détecter des comportements anormaux.
- Entitlement Management (CIEM). Cette mesure permet la vérification des droits d'accès associés avec des rôles et des utilisateurs dans le cloud et ainsi de proposer des optimisations afin de promouvoir une approche moindre privilège.

2. **Mesures de protection des données « Policy Enforcement Point** » a pour objectif d'appliquer des mesures de protection associées à la politique de protection, incluant le blocage d'accès ou de partage de données sous l'instruction du PDP (Policy Decision Point). Ces mesures de protection peuvent être regroupées sous les catégories d'encodage, de prévention contre la fuite des données et de Information Rights Management.

   - **L'encodage.** Cette mesure a pour effet de transformer les valeurs et éventuellement le format des données afin de les désensibiliser et ainsi protéger les informations initiales associées avec le scénario d'usage. Ce type de protection est typiquement appliqué aux données au repos « at rest » afin d'assurer leur confidentialité et la protection de la vie privée.

     Le chiffrement est la méthode d'encodage la plus répandue et peut être appliqué à différents niveaux, par exemple la base de données ou par l'application. Alternativement, des nouvelles techniques d'encodage telles que la pseudo-anonymisation, la tokenization, l'anonymisation et le masquage (Data masking) peuvent offrir des capacités d'encodage plus adapté pour des règles granulaires au niveau données, sans exiger des transformations dans le schéma des bases de données.
     Toutefois l'encodage ne pourra être employé sans la mise en place des mécanismes de gestion des clés incluant les processus de génération, distribution et stockage ainsi que la rotation et la révocation des clés.

   - **La prévention contre la fuite des données** (« Data Loss Prevention ») permet de sécuriser le partage des données sensibles avec des tiers ainsi que d'empêcher des envois de données vers l'extérieur qui ne sont pas compatibles avec les scénarios d'usage habituels. La réflexion typique des équipes IT est de considérer que les données sensibles ne doivent jamais sortir du périmètre de l'entreprise. Toutefois cette option n'est souvent pas viable là où les différents processus métiers nécessitent le partage de données sensibles avec l'extérieur afin de supporter des scénarios d'usage bien spécifiques.

   - **Information Rights Management (IRM)**. Ainsi, au-delà de la protection contre les fuites périmétriques, des fonctions de protection doivent s'intégrer dans la donnée elle-même pour avoir la possibilité de voyager avec cette donnée et assurer sa protection persistante tout au long de son cycle de vie, même hors du périmètre de l'entreprise. Afin de supporter ces besoins, les mesures de protection peuvent être étendues et inclure de l'"Information Rights Management (IRM)" ainsi que du «Data In Use, Loss Prevention on Endpoint ».

6. Illustration du modèle proposé à deux scénarios d'usage

Nous proposons d'illustrer l'application de ce modèle à travers deux scénarios d'usage typiques des données. Le premier de ces scénarios concerne l'ingestion, le stockage et l'analyse des données par les métiers.

SCENARIO (1) : ANALYSE DES DONNEES

Pour ce premier scénario d'usage nous considérons la consommation des données sensibles dans un environnement Big Data. Dans ce cas, nous considérons trois environnements qui sont employés afin de faciliter la collecte, le stockage et la mise à disposition ainsi que l'analyse des données.

1. **Plateforme d'ingestion des données (Pipeline d'ingestion)**

   Le pipeline d'ingestion des données gérera la collecte de données à partir d'une variété de sources structurées, semi-structurées et non structurées dans le but de consolider toutes ces données dans un hub de données central.

   La collecte peut être effectuée via un processus par lots planifiés réguliers ou via un streaming en temps réel. Les données entrantes, avant d'être stockées dans le hub de données centralisé, seront inventoriées, cataloguées et classées.

2. **Plateforme de Stockage des données (Centre/ Lac de données)**

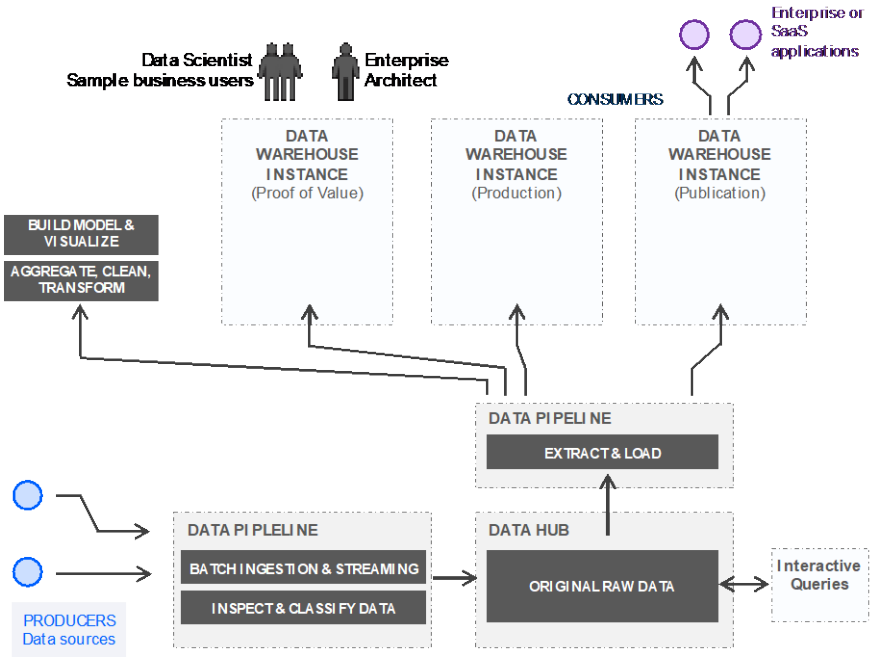
   Le Data Hub sera le point de stockage central où les différentes sources des données diffuseront leurs contenus. Les données consolidées dans le Hub sont accessibles par diverses méthodes qui ne sont pas toujours souhaitables, notamment les accès directs par les administrateurs de la plateforme.

   Les accès les plus fréquents sur un Data Hub restent limités au chargement des entrepôts de données spécifiques à l'application et la consommation des données par les applications BI et AI qui reçoivent des données en batch ou en streaming du Hub.

3. **Plateforme de consommation des données (entrepôts de données)**

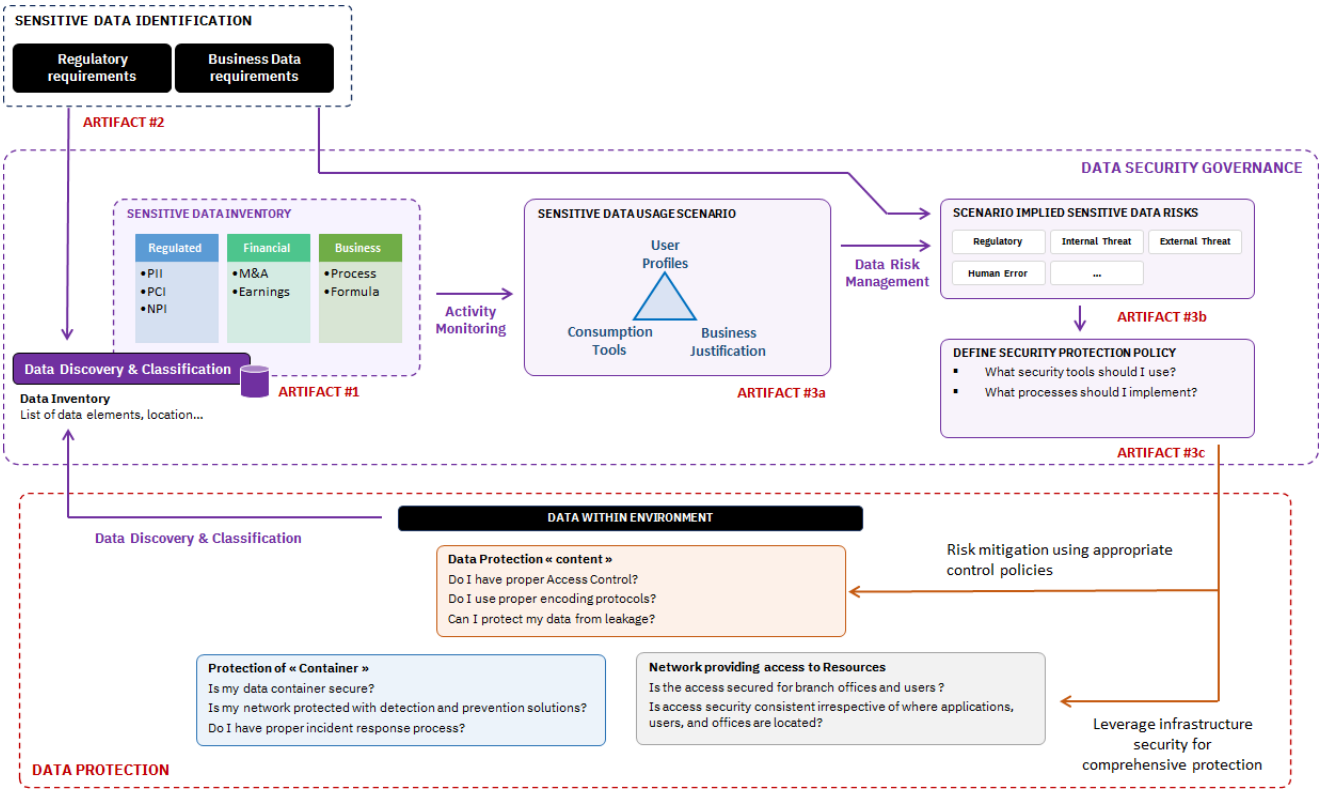
   Au niveau de l'entrepôt de données, seule une sélection spécifique de données est ingérée et pour des traitements métiers spécifiques. La donnée est stockée dans un objectif bien déterminé et pour une finalité bien gouvernée par les métiers.

   Contrairement à un Hub de données, les entrepôts bénéficient d'un schéma bien documenté permettant une consommation agile et optimale des données par les analystes et les applications.



Adaptation au modèle de sécurité proposé

Afin d'assurer une protection des données et de la vie privée à la conception de cette plateforme de données nous adapterons cette architecture au modèle de sécurité proposé.



Artefact 1 & 2 – Identification des données sensibles

La première activité à prendre en considération sera l'identification et la découverte des données sensibles. Ci-dessous une exemple une liste non exhaustive de type de données que nous pouvons trouver :

Données reglementées	GDPR/ PII	Nom, Prénom
		Date de naissance
		Adresse, Téléphone
		Sécurité Sociale
	PCI/ DSS	IBAN
		Cartes Crédits
	HIPPA	Pathologie
		Procédure médicale
Données de Gestion de l'Entreprise	Chiffre d'affaire	
	Revenus	
	Aquisition	
	Gestion d'actifs	
Données Metiers d'Entreprise	Procédure de Fabrication	
	Formule chimique	
	Brevet	
	Secret professionnel	

Artefact 3 – Matrice d'usage et de protection des données

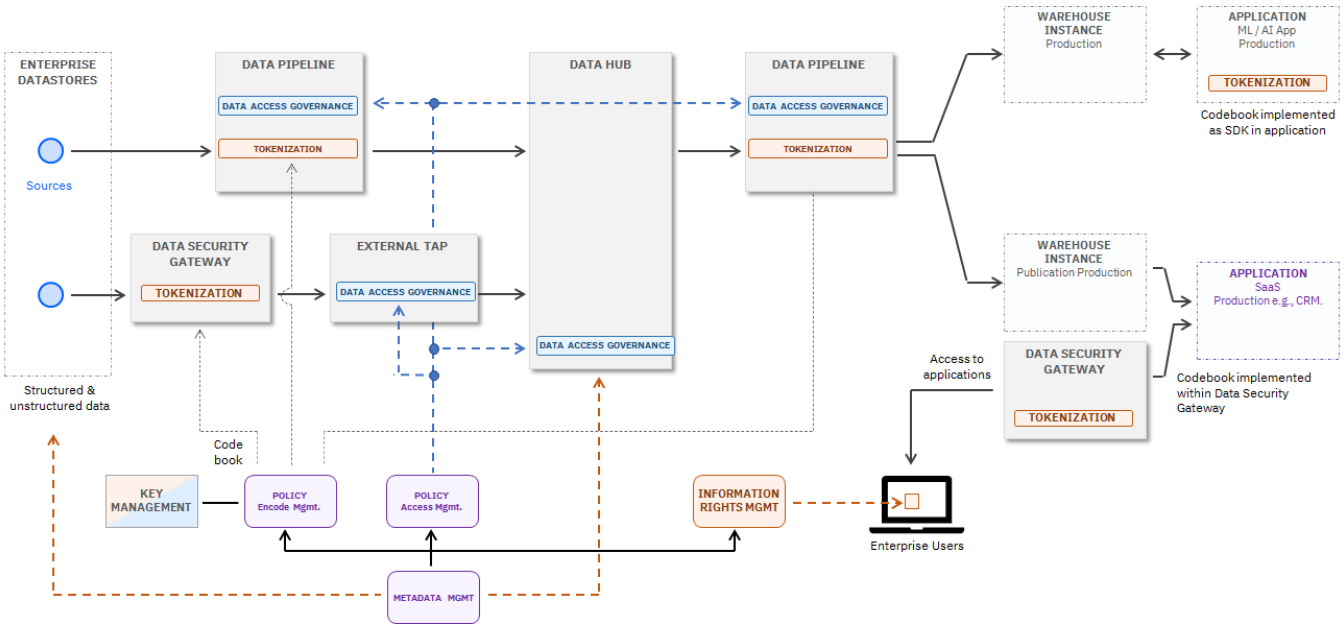
Artefact 3a – Identification des scénarios d'usage

Artefact 3b – Identification des risques induits

Artefact 3c – Identification des fonctions de Sécurité requises

Artifact 3a			Artifact 3b	Artifact 3c		Data Protection
Profile	**Platform access**	**Business Justification**	**Risk**	**Controls**	**Governance Contols**	**Tooling**
Administrator	Data Hub	Platfrom administration	Access to sensitive data	Data Level Access Control	Activity Monitoring	Guardium
	Data Warehouse	Platfrom administration	Access to sensitive data	Data Level Access Control	Activity Monitoring	Guardium
Data Analyst	Data Hub	Business usage	Abuse of sensitive data	Data Tokenization	Activity Monitoring	Protegrity
	Data Warehouse	Business usage	Abuse of sensitive data	Data Tokenization	Activity Monitoring	Protegrity
	Endpoint	Business Report	External sharing	Data Right Management	Activity Monitoring	AIP / VERA
Developper	Endpoint	Application enhancement	Abuse of sensitive data	Data Anonymization	N/A	Protegrity

Data Protection – Architecture de protection des données



SCENARIO (2) : MICROSOFT 365

Pour ce deuxième scénario d'usage nous considérons la consommation des données sensibles dans un environnement M365.

Malgré tous les contrôles de sécurité employés au niveau des environnements, le maillon faible de la chaîne reste l'humain. Afin de consommer les données, des utilisateurs auront besoin d'y accéder pour effectuer leurs activités métiers. Ainsi ces utilisateurs pourront représenter des acteurs de menaces contre la protection des données dans le cloud soit par étourderie où même de manière intentionnelle.

Considérons ce cas d'usage autour de la consommation des données dans les services SaaS tels que les services Microsoft Office 365. Cette plateforme de collaboration permet aux utilisateurs plusieurs services autour de la donnée incluant :

- Création/ modification des documents sensibles
- Stockage en local ou dans le cloud
- Partage par mail, Visio, Chat ou à travers l'environnement de stockage (OneDrive)

Ainsi pour assurer la protection des données sensibles dans cet environnement nous retournons sur notre approche basée sur l'usage afin d'identifier les menaces et proposer les contrôles appropriés.

Dans une première étape, il est nécessaire d'énumérer les différents types de données considérées sensibles pour l'entreprise qui devront être protégés (Artefact #**1**). Nous distinguons trois catégories de données : les données règlementées, les données spécifiques aux différents secteurs métiers et les données spécifiques à l'Entreprise.

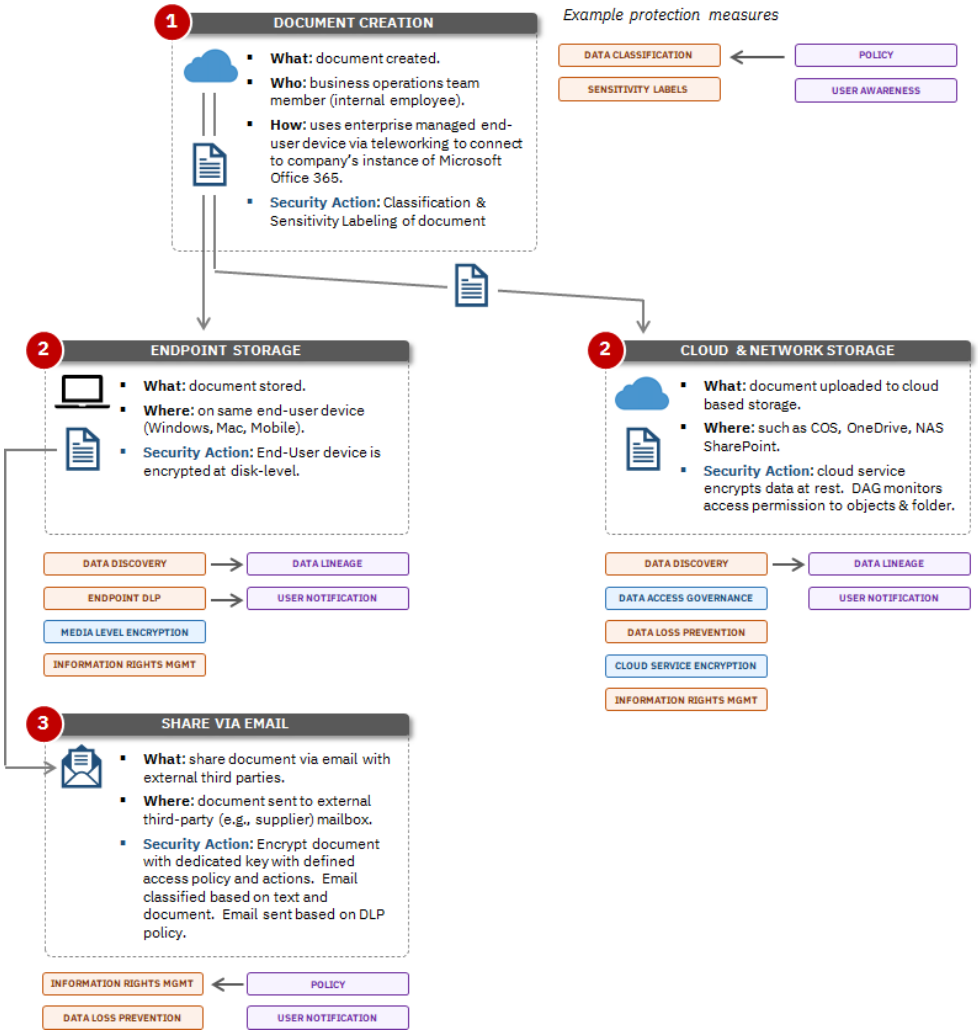
L'étape suivante revient à la définition des contextes d'usages associés avec les données sensibles identifiées dans Artefact **#1** dans le but de les protéger. Par exemple nous proposons la liste d'usages suivante qui pourra avoir un impact sur la protection des données :

- Envois de Données sensibles vers l'extérieur à travers mail/ chat

- Divulgation de données à travers un partage interne
- Transfert des données à caractère personnel en dehors de l'Union Européen.

Une fois les usages identifiés (Artefact #3a), nous pourrons commencer à regarder les risques ainsi que les contrôles fonctionnels appropriés pour chacun des usages (Artefacts #3b & 3c). Par exemple pour le premier scénario d'usage sur le partage extérieur, les données sortent de l'environnement gouverné par la sécurité et sont mises à disposition du récepteur qui pourra les consommer comme il lui convient. Delà il faut penser à des contrôles qui peuvent voyager avec la donnée, comme les solutions « Information Rights Management ». Dans ce cas même en dehors du périmètre de l'entreprise, la politique de sécurité reste toujours attachée au document envoyé et les activités du récepteur seront contrôlées et tracées par la solution pour assurer la conformité avec la finalité d'usage prévue pour cette donnée.

Nous avons illustré dans le diagramme ci-dessus l'association potentielle entre les cas d'usage métier (création, stockage et partage avec des tiers) et les mesures de protection.



D'autre part sur les partages en interne, une approche complètement différente pourra être considérée. Les données restant dans le périmètre de l'entreprise, des contrôles de type « Information Rights Management » pourront être mis en place pour assurer la protection des données. Ainsi les droits de

lecture/ écriture sont revus sur tous les environnements de partage pour valider les droits avec les propriétaires de la donnée. De plus nous pouvons remédier les problèmes de groupes imbriqués ou des accès directs afin d'assurer une gouvernance robuste des droits d'accès aux données.

7. Conclusion

Le modèle présenté dans ce document propose une approche de protection des données tout au long de leur cycle de vie. Il prend en compte les risques associés avec les nouveaux usages des données tels que de l'Analytics et du Cloud.

A travers sa représentation de la gouvernance et des mesures de protection, ce modèle facilite la collaboration et l'échange entre les différentes équipes au sein de l'entreprise (ex : métier, risques et conformité, sécurité, production informatique). Le cadre fourni par ce modèle est souple et favorise le passage à l'échelle des nouveaux cas métiers dans un cadre sécurisé et contextualisé, se basant sur l'usage de la donnée associé à ces nouveaux cas. De plus, il promeut une approche globale et cohérente entre les fonctionnalités de sécurité requises par ces différents cas d'usages, évitant ainsi une approche par silo des différentes solutions techniques.

La consommation et le traitement des données (ex : identifier des tendances ou des corrélations) au travers des nouveaux services de type Cloud (PaaS, SaaS) introduit des nouveaux risques (ex : souveraineté digitale et l'accès éventuel par les administrateurs du CSP). Cependant, pour les données à caractère personnel, les besoins de traitement des données en clair impliquent également de nouveaux risques tels que l'identification des attributs d'un individu par inférence (à travers d'autres attributs).

Parallèlement, nous constatons que le marché évolue vers un modèle de protection de la vie privée et de sécurité centrée sur les données. Dans un tel modèle, le contexte d'usage est primordial afin d'identifier les risques associés aux usages et ainsi proposer des mécanismes de protection adéquats. Cette approche est capable d'aborder à la fois les usages s'appuyant sur de nouvelles architectures conçues, par exemple, sur le Cloud ou des infrastructures traditionnelles.

Afin de commencer à appliquer ce modèle, nous proposons d'adopter la démarche suivante :

Mettre en place cadre Organisationnel de l'entreprise…

- Identifier un rôle de référent des équipes métier (et de sécurité) vis-à-vis des usages des données au sein de l'entité métier, qui sera intégré dans l'équipe de Data Management des métiers.
- Intégrer les activités décrites dans ce document, incluant la modélisation des mesures de protection, dans la phase « conception » des projets métiers et IT.

Construire des cas d'usages métier – approche incrémentale / itérative du modèle

- Choisir un projet comme pilote afin de comprendre les scénarios de l'usage métier des données ainsi qu'identifier les profils d'utilisateurs
- Identifier les risques, ainsi que les mesures de protection associées.
- Etablir un inventaire des données pendant la phase de développement d'une application voire la transformation digitale d'un processus métier.

Mise en cohérence -- Intégration des mesures de protection dans l'infrastructure

- S'appuyer sur une équipe de « Enterprise Security Architecture » afin d'aligner et d'agréger les besoins de mesures de protection identifiées dans les différents projets.
- Effectuer le choix et l'alignement des solutions techniques.

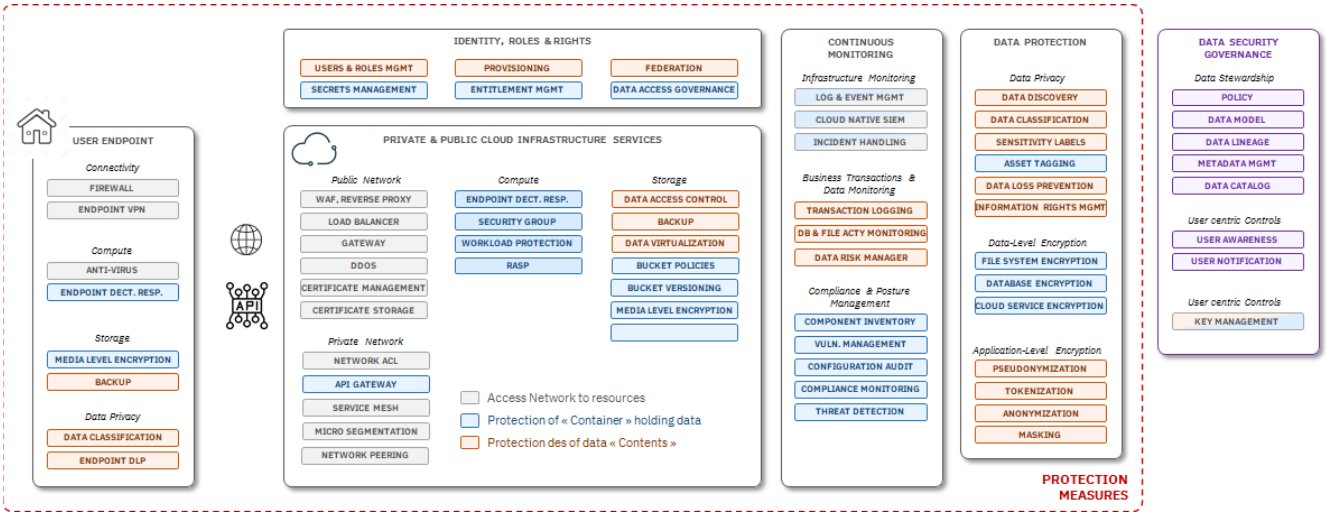
Le modèle lui-même doit aussi être adaptable face à l'évolution des réglementations ainsi qu'à l'émergence de nouvelles technologies de consommation des données.

Les réglementations sur les données personnelles associées à la consolidation et au traitement des données personnelles continueront de se renforcer au fur à mesure dans les différents pays.

Ces évolutions peuvent présenter des limites aux possibilités d'analyse des données par les métiers. De ce point de vue, le modèle pourra être enrichi à la suite de l'expérimentation de nouvelles technologies. Par exemple, les solutions telles que FHE (Fully Homomorphic Encryption) et Differential Privacy permettent aux métiers d'explorer des nouveaux cas d'usage sur la base des données consolidées entre plusieurs pays ou entreprises qui ne sont pas possibles aujourd'hui avec la réglementation en place (ex : identification du blanchiment des capitaux avec l'ACPR : https://acpr.banque-france.fr/autoriser/fintech-et-innovation/experimentation-sur-la-mutualisation-de-donnees-en-lcb-ft).

8. Glossaire

Définition des mesures de protection illustrées dans le cadre ci-dessus.



Continuous Monitoring

Log & Event Management	▪ *Log Management is he process for generating, transmitting, storing, analyzing, and disposing of log data [NIST SP 800-92].* ▪ Collect event logs from cloud-native services and the management plane as well as services deployed on-top of cloud. ▪ Can perform threat hunting using collected logs.
Cloud Native SIEM	▪ *Application that provides the ability to gather security data from information system components and present that data as actionable information via a single interface [NIST SP 800-128].* ▪ *Event Correlation involves finding relationships between two or more log entries [NIST SP 800-92].* ▪ Correlate events and alerts using combination of rules and anomaly-based detection to identify potential security incidents. ▪ Threat Intelligence used to enhance detection capabilities.
Incident Handling	▪ *Incident Handling involves the mitigation of violations of security policies and recommended practices [NIST SP 800-61r2].* ▪ Once the potential security incident has been identified, it is enriched, with threat intelligence and contextual information, and qualified. ▪ Investigation and response are subsequently performed by level 2 & 3 teams.
Transaction Logging	▪ *A transaction is a discrete event between a user and a system that supports a business or programmatic purpose [NIST SP 800-63].* ▪ Monitor who accesses what (column, row, table, file) and when. ▪ For structured datastores, solutions typically log contents of INSERT actions as well as SELECT queries to identify, alert, and potentially block actions not compliant with policy.

Database & File Activity Monitoring	▪ *Database activity monitoring (DAM) can be used to support the ability to identify and report on fraudulent, illegal or other undesirable behavior, with minimal impact on user operations and productivity. The tools, which have evolved from basic analysis of user activity in and around relational database management systems (RDBMSs) to encompass a more comprehensive set of capabilities, such as discovery and classification, vulnerability management, application-level analysis, intrusion prevention, support for unstructured data security, identity and access management integration, and risk management support [GARTNER].* ▪ *File Activity Monitoring discovers the sensitive data on your servers; classifies content using pre-defined or user defined definitions; configures rules and policies about data access, and actions to be taken when rules are met [IBM].*
Data Risk Manager	▪ *Provides executives and their teams a business-consumable data risk control center that helps to uncover, analyze, and visualize data-related business risks so they can take action to protect their business [IBM].* ▪ Visualize risks associated with sensitive data based on aggregated view of policy violations, vulnerabilities, and threats associated with supporting services and assets.
Component Inventory	▪ *System components are discrete, identifiable information technology assets that include hardware, software, and firmware. The System Component Inventory accurately reflects the system and includes all components in the system [NIST SP 800-53r3].* ▪ Maintain an exhaustive list of provisioning assets including nodes/pods/clusters in CaaS platform.
Vulnerability Management	▪ *Vulnerability Management is an ISCM capability that identifies vulnerabilities [Common Vulnerabilities and Exposures (CVEs)] on devices that are likely to be used by attackers to compromise a device and use it as a platform from which to extend compromise to the network [NISTIR 8011 Vol. 1].* ▪ *A Vulnerability is a weakness in an information system, system security procedures, internal controls, or implementation that could be exploited or triggered by a threat source [NIST SP 800-30].* ▪ Identify presence of vulnerabilities (CVE) within cloud-based services as well as services deployed on-top of cloud.
Configuration Audit	▪ *Configuration Settings Management is an ISCM capability that identifies configuration settings (Common Configuration Enumerations [CCEs]) on devices that are likely to be used by attackers to compromise a device and use it as a platform from which to extend compromise to the network [NIST IR 8011-1].* ▪ Verifies misconfiguration of cloud services.
Compliance Monitoring	▪ *Compliance monitoring verifies that required risk response measures are implemented. t also verifies that security and privacy requirements are satisfied. [NIST SP 800-53r3].* ▪ Involves the proactive validation that internal controls are in place and functioning as expected. ▪ Reviews evolving compliancy of cloud services with respect to security frameworks such as CIS Safeguards & Benchmarks.
Threat Detection	▪ *A Threat is any circumstance or event with the potential to adversely impact organizational operations, organizational assets, individuals, other organizations, or the Nation through a system via unauthorized access, destruction, disclosure, modification of information, and/or denial of service [NIST SP 800-30].* ▪ Leverage detection techniques such as anomaly detection, UEBA, and network alerts which are offered posture management solutions.

Data Protection

Data Discovery	▪ *The process of analyzing the type, quality, accessibility, and location of data in all available data repositories. It's critical for determining the current state of a data environment, especially when a recent and accurate data dictionary doesn't exist [FORRESTER].* ▪ Involves scanning the environment to determine the presence of datastores and the exploration of the data objects contained in these stores.
Data Classification	▪ *Data discovery and classification represents the capability to provide visibility into where sensitive data is located, identify what this sensitive data is and why it's considered sensitive, and tag or label this data based on its level of sensitivity [FORRESTER].* ▪ Leverage predefined or custom policy rules based on such as keywords, AI, and patterns to identify the presence of specific data types and values within the datastores. ▪ Object labels can subsequently be applied to indicate the determined data sensitivity.
Sensitivity Labels	▪ *Sensitivity labels are tags that allow organizations to categorize data. Sensitivity labels applied on an asset travel with the data. [AZURE PURVIEW].* ▪ Apply data sensitivity labels to such as files and columns within structured data stores.
Asset Tagging	▪ *Can assign metadata to resources in the form of tags. Each tag is a label consisting of a user-defined key and value. Tags can help you manage, identify, organize, search for, and filter resources. You can create tags to categorize resources by purpose, owner, environment, or other criteria [AWS TAGGING RESOURCES].* ▪ Tag assets provisioned within the cloud to indicate such as environment (e.g., prod, staging) and compliance requirements (e.g., HIPAA).
Data Loss Prevention	▪ *A capability that detects and prevents violations to corporate policies regarding the use, storage, and transmission of sensitive data. Its purpose is to enforce policies to prevent unwanted dissemination of sensitive information [FORRESTER].* ▪ Inspect the contents of data objects at rest, in use, and in motion to protect against data leaving the environment boundary. ▪ DLP can be implemented either on the end user device or on network gateways (e.g., mail, web). ▪ Actions can include alert, relocate, tag, and encrypt.
Information Rights Management	▪ *Information Rights Management (IRM) is a form of IT security technology used to protect documents containing sensitive information from unauthorized access. Unlike traditional Digital Rights Management (DRM) that applies to mass-produced media like songs and movies, IRM applies to documents, spreadsheets, and presentations created by individuals. IRM protects files from unauthorized copying, viewing, printing, forwarding, deleting, and editing [MCAFEE].* ▪ *Information Rights Management (IRM) helps organizations enforce information policies by allowing the author of electronic documents or emails to choose how recipients can use and share the information. This helps protect a company's intellectual property and helps prevent critical or damaging business information from being disclosed accidentally. IRM allows you to set policies over who can open, copy, print, or forward information [MICROSOFT].* ▪ Enable secure sharing of documents internally and externally. Security applied to individual files allowing control of who accesses the document and the actions that can be performed. Access can be revoked remotely as necessary.

File System Encryption	▪ Transparent data encryption mechanism that allows granular access control and data access logging for any storage resource (Folder, Volume, Cloud…) ▪ Local agent can monitor OS authenticated users and leverages centralized user policy/ KMS to provide access to protected resources in filesystem. ▪ Access Logging allows better investigation capabilities to identify and stop threats faster.
Database Encryption	▪ Also known as "Transparent Database Encryption, ensures all database is encrypted written in the file system is encrypted by the database server prior to right action. ▪ On consumption, the database application reads the encrypted data and executes data de-encryption in memory in real time to provide access to data in clear. ▪ The content of the database is protected by a symmetric key however KMS integration should be considered to protect the encryption key in the DB Server.
Cloud Service Encryption	▪ *Enforce encryption at rest by ensure that the only way to store data is by using encryption. KMS integrates seamlessly with many AWS services to make it easier for you to encrypt all your data at rest [AWS WELL-ARCHITECTED SECURITY PILLAR].* ▪ Data associated with different storage services can be encrypted using a data encryption key (DEK) associated with that service. ▪ The data encryption key is typically protected by wrapping it with a customer managed key.
Pseudonymization	▪ *Pseudonymization means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person [GDPR ARTICLE 4]* ▪ Performs de-identification of Personally Identifiable Information (PII). In this case the value of a data element is replaced systematically by an artificial identifier.
Tokenization	▪ *Tokenization refers to a process by which a piece of sensitive data, such as a credit card number, is replaced by a surrogate value known as a token. The security of a tokenization approach depends on the security of the sensitive values and the algorithm and process used to create the surrogate value and map it back to the original value [GARTNER].* ▪ Involves the substitution of a sensitive data element by a non-sensitive value (token) without altering its type or length with enables processing. The non-sensitive value is used as a token for the tokenization system to map back to the original value. ▪ This is considered an example of pseudo anonymization.
Anonymization	▪ *A method of de-identification that removes all personally identifiable information from a data set to the extent that makes the possibility of re-identification of the data negligible [FORRESTER].* ▪ Involves the modification of personal data such that the data subject can no longer be identified directly or indirectly.
Masking	▪ *The process of obfuscating sensitive data in production and non-production environments to support data security and compliance requirements [FORRESTER].* ▪ Involves the modification of the original data element such that its value is no longer sensitive, while preserving the structure of the data. ▪ This can be realized using several techniques including tokenization and data obfuscation. .

Data Security Governance

Policy	▪ *Codify principles and management intent into fundamental rules governing the creation, acquisition, sharing, processing, storage, integrity, security, quality and use of data and information. Policy is global and supports data standards [DMBOK].*
Data Model	▪ *The Data Model describes the database structure, its relationships and the constraints that determine how data is stored (even physically), encoded and accessed (e.g., Entity Relationship Model).* ▪ *The Semantic Data Model is a method of organizing data that reflects the basic meaning of data items and the relationships among them. This organization makes it easier to develop application programs and to maintain the consistency of data when it is updated [DMBOK].*
Data Lineage	▪ *Provide an IT mapping that ensure the data life cycle visualization through flows description.* ▪ *Provide visibility of the data origin, data flows / movement between source and destination as well as transformations that have been applied to the data.* ▪ *Valuable for security in ensuring the appropriate classification of data as it moves as well as being combined and transformed [DMBOK].*
Metadata Management	▪ *Metadata is information that describes various facets of an information asset to improve its usability throughout its life cycle: data use or data own, data sensitivity and business definition.* ▪ *Metadata contains security, privacy related context or will be leveraged by security, and privacy related activities. Metadata thus turns information into an asset.* ▪ *Generally speaking, the more valuable the information asset, the more critical it is to manage the metadata about it, because it is the metadata definition that provides understanding that unlocks the value of data [DMBOK].*
Data Catalog	▪ *A data catalog is a technology capability that is used to manage an inventory of heterogeneous and distributed data assets through the discovery, organization and description of the enterprise datasets.* ▪ *A data catalog consists of an Inventory of data sources, a Technical Dictionary and a Business Glossary.* ▪ *It provides context to help data architect, developers, data analysts, data engineers, data scientists, data stewards and other data consumers to locate a relevant dataset and understand what it means, in order to determine and extract business value from it.* ▪ *[DMBOK, DUBLN CORE METADATA INITIATIVE (DCMI) to create metadata model].*
User Awareness	▪ *A learning process that sets the stage for training by changing individual and organizational attitudes to realize the importance of security and the adverse consequences of its failure [NIST SP 800-16].* ▪ Regularly educate and test users to ensure their understanding and application of data security and privacy policies. ▪ User notifications, associated with such as DLP, based upon policy violations also contribute to user awareness.
User Notification	▪ Alert user that their action will result (or has resulted) in a violation of security or privacy policies.
Key Management	▪ *The activities involving the handling of cryptographic keys and other related key information during the entire lifecycle of the keys, including their generation, storage, establishment, entry and output, use, and destruction [NIST SP 800-57 Part 1 r5].*

	▪ *Manage keys through their lifecycle including creation, association with a cloud service, rotation, and revocation.* ▪ *Keys generated through cloud-based key management system typically used to wrap the data encryption keys employed by the different cloud services.* ▪ *Secure storage of keys delivered through an HSM which can be either shared or dedicated.*

Identity, Roles ad Rights

Identity Management	▪ *Identity management is an activity within the identity and access management function that concerns the governance and administration of a unique digital representation of a user, including all associated attributes and entitlements [GARTNER].*
Provisioning	▪ *User provisioning or account provisioning technology creates, modifies, disables and deletes user accounts and their profiles across IT infrastructure and business applications. Provisioning tools use approaches such as cloning, roles and business rules so that businesses can automate onboarding, offboarding and other administration workforce processes (for example, new hires, transfers, promotions and terminations). Provisioning tools also automatically aggregate and correlate identity data from HR, CRM, email systems and other "identity stores [GARTNER].*
Federate	▪ *Federated identities are those which enable users to have a single identity stored in an organization's central identity provider [AWS FEDERATED IDENTITIES].*
Secrets Management	▪ *Helps you protect secrets needed to access your applications, services, and IT resources. The service enables you to easily rotate, manage, and retrieve database credentials, API keys, and other secrets throughout their lifecycle. Text [AWS SECRETS MANAGER].*
Entitlement Management	▪ *Entitlement management is technology that grants, resolves, enforces, revokes and administers fine-grained access entitlements (also referred to as "authorizations," "privileges", "access rights," "permissions" and/or "rules"). Its purpose is to execute IT access policies to structured/unstructured data, devices and services [GARTNER].*
Data Access Governance	▪ *Data-Centric directory services governance platform. Enhances access rights provided by Directory services to folders or other containers based on the sensitivity of the data stored within.* ▪ *DAG Solutions scan managed storage environments for sensitive data, identify access rights provided through directory services and remediates authorizations which are non-compliant with policies.*

Private & Public Cloud Infrastructure Services

WAF, Reverse Proxy	▪ *A ''web application firewall (WAF)'' is an application firewall for HTTP applications. It applies a set of rules to an HTTP conversation. Generally, these rules cover common attacks such as Cross-site Scripting (XSS) and SQL Injection. While proxies generally protect clients, WAFs protect servers. A WAF is deployed to protect a specific web application or set of web applications. A WAF can be considered a reverse proxy [OWASP].* ▪ Analyze protocol and network flows arriving and leaving application to detect breaches and stop attacks.

Load Balancer	▪ *A load balancer serves as the single point of contact for clients. It automatically distributes incoming traffic across multiple targets, such as EC2 instances, containers, and IP addresses, in one or more Availability Zones [AMAZON ELASTIC LOAD BALANCING].*
Security Gateway	▪ *An internetwork gateway that separates trusted (or relatively more trusted) hosts on one side from untrusted (or less trusted) hosts on the other side [IETF RFC 4949 GLOSSARY].* ▪ *Can be represented by such as a VPN gateway or firewall.*
Anti-DDoS	▪ *A Distributed Denial of Service (DDoS) is a technique that uses numerous hosts to perform the attack [NIST IR 7711].*
Certificate Management	▪ *Process whereby certificates are generated, stored, protected, transferred, loaded, used, and destroyed [CNSSI 4009-2015].*
Certificate Storage	▪ *Provides storage as mentioned within above Certificate Management definition*
Network ACL	▪ *Acts as a firewall for controlling traffic in and out of a subnet. You can associate multiple subnets with a single network ACL, but a subnet can be associated with only one network ACL at a time [AWS NACL].*
API Gateway	▪ *An API gateway is an API management tool that sits between a client and a collection of backend services. An API gateway acts as a reverse proxy to accept all application programming interface (API) calls, aggregate the various services required to fulfill them, and return the appropriate result [REDHAT].*
Service Mesh	▪ *A service mesh is a dedicated infrastructure layer that you can add to your applications. It allows you to transparently add capabilities like observability, traffic management, and security, without adding them to your own code [ISTIO Open-Source SERVICE MESH].*
Micro Segmentation	▪ *An approach to network security where access to network resources is granted by defined policy, using established relationships between identities, and not simply placement within the network topology [FORRESTER].*
Network Peering	▪ *A [VPC] peering connection is a networking connection between two VPCs that enables you to route traffic between them using private IPv4 addresses or IPv6 addresses. Instances in either VPC can communicate with each other as if they are within the same network. Can create a VPC peering connection between your own VPCs, or with a VPC in another account. The VPCs can be in different regions, also known as an inter-region VPC peering connection. [AMAZON VPC PEERING].*
Endpoint Detection & Response	▪ *EDR (Endpoint Detection and Response) covers some more advanced capabilities like detecting and investigating security incidents, and ability to remediate endpoints to pre-infection state [SANS].* ▪ Solution deployed on user endpoints (laptop, desktop, mobile) and servers. ▪ Typically provides capabilities to detect, investigate, and response to threats as well as facilitate threat hunting.
Security Group	▪ *A security group acts as a virtual firewall for your EC2 instances to control incoming and outgoing traffic. Inbound rules control the incoming traffic to your instance, and outbound rules control the outgoing traffic from your instance [AMAZON EC2 SECUIRTY GROUPS].* ▪ Stateful firewall associated with cloud compute instances which filters incoming and outgoing traffic based upon the defined ruleset.
Workload Protection	▪ *Workload-centric security solution that targets the unique protection requirements" of workloads in modern enterprise environments [GARTNER].* ▪ Continuously monitor container behavior to detect suspicious activities.

	▪ Detect containers drifting from their image that may indicate a compromission.
RASP	▪ *Runtime application self-protection (RASP) is a security technology that is built or linked into an application or application runtime environment and is capable of controlling application execution and detecting and preventing real-time attacks [GARTNER].* ▪ Runtime Application Self Protection (RASP) complements perimeter-based protection and is embedded within the application's runtime environment. ▪ A RASP solution analyzes how the requests and data are processed within the application.
Data Access Control	▪ Granular object level access control policies managing access to sensitive data. Access to objects and columns containing sensitive data is monitored, tracked and blocked if not identified. ▪ The solution is positioned on the storage server monitoring and controlling queries at the system level providing segregation of duties with platform administrators.
Backup	▪ *A copy of files and programs made to facilitate recovery if necessary [NIST SP 800-34r1].* ▪ Provide data backup and restore services for block, database, file, and object-based datastores. Typically delivers back services to both cloud and on-premises-based resources.
Data Virtualization	▪ *The integration and transformation of data in real time or near real time from disparate data sources in multi-cloud and hybrid cloud, to support business intelligence, reporting, analytics, and other workloads [FORRESTER].* ▪ Provides a holistic view of enterprise data across a wide diversity of underlying sources without having to consolidate and normalize the data into a central repository.
Bucket Policies	▪ *A bucket policy is a resource-based AWS Identity and Access Management (IAM) policy. Add a bucket policy to a bucket to grant other accounts or IAM users access permissions for the bucket and the objects in it. Object permissions apply only to the objects that the bucket owner creates [AMAZON S3].* ▪ These are resource based IAM policies that can be used to grant permission for object storage. A bucket policy applies to a bucket and the objects it contains.
Bucket Versioning	▪ *Versioning is a means of keeping multiple variants of an object in the same bucket. You can use the S3 Versioning feature to preserve, retrieve, and restore every version of every object stored in your buckets. With versioning you can recover more easily from both unintended user actions and application failures [AMAZON S3 BUCKET VERSIONING].* ▪ Retain and restore different versions of every object contained within a bucket. Supports recovery from accidental object deletion or overwrite.
Media Level Encryption	▪ Hardware level encryption, Ensuring protection of an Endpoint or physical server. ▪ Complete system is encrypted on shutdown and decrypted during system boot is user is authorized to boot the system.

Définition des rôles

Data Protection Officer (DPO)	▪ The primary role of the data protection officer (DPO) is to ensure that her organization processes the personal data of its staff, customers, providers or any other individuals (also referred to as data subjects) in compliance with the applicable data protection rules.
Chief Data Officer (CDO)	▪ Senior Executive responsible for the utilization and governance of his organization data. His primary objective is to benefit the most from the organization main assets: it's data.

Data Stewards	▪ The data stewards manage the data assets for the best interests of the organization. Their main activities involve creating and managing metadata, documenting rules and standards, managing data quality and executing operation data governances. They can be differentiated by their place within the organization and focus of their work (CDO, DG Council, Data/ Domain owner, Technical …).

9. Références

1.	Information: the currency of the digital age Former HP CEO Carly Fiorina http://www.hp.com/hpinfo/execteam/speeches/fiorina/04openworld.html
2.	Inside BNP Paribas' Digital Banking Innovation: Cloud, Data, AI BNP Paribas Global CIO Bernard Gavgani https://www.delphix.com/blog/inside-bnp-paribas-digital-banking-innovation-cloud-data-ai