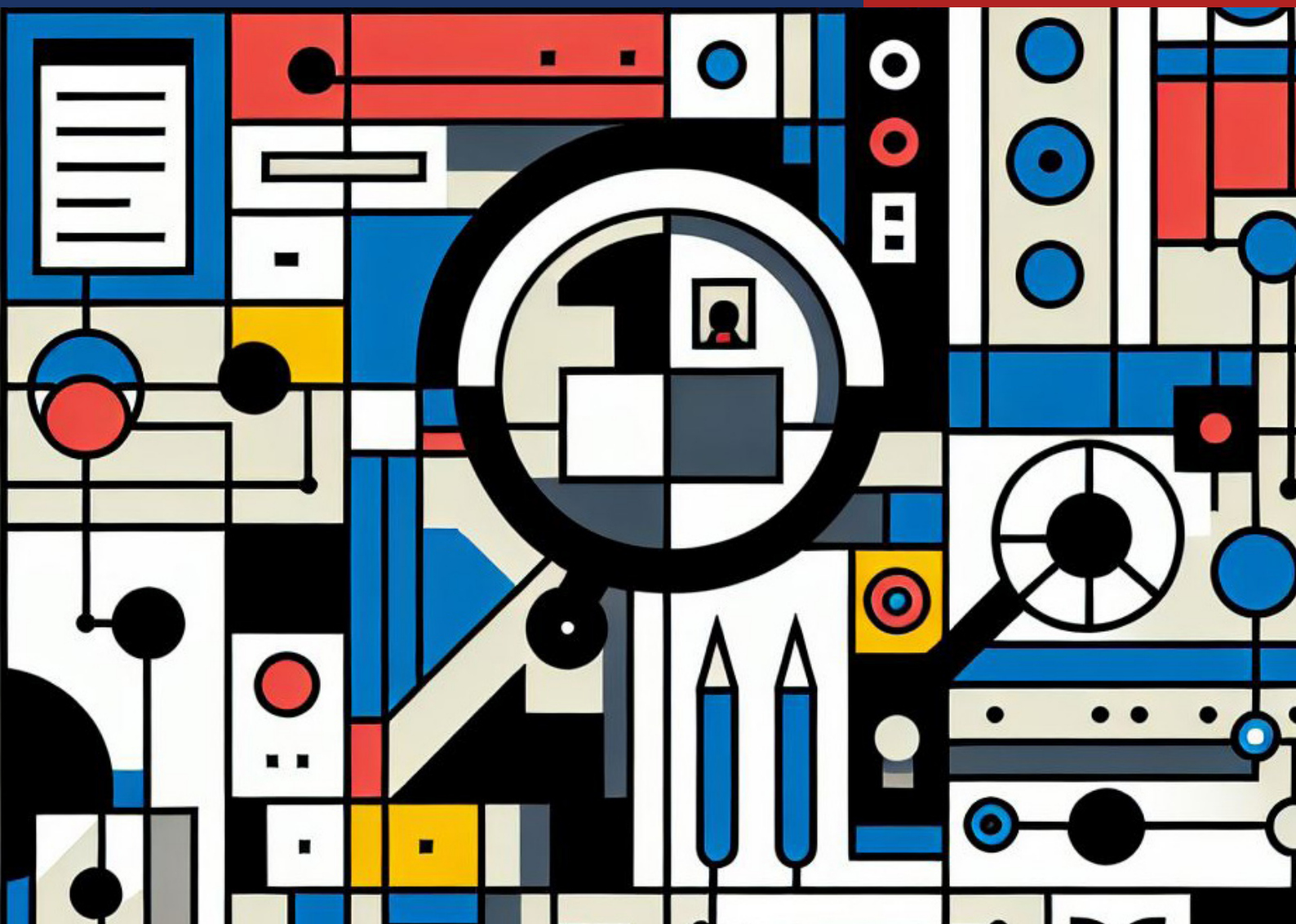
CNPD

COMMISSION NATIONALE POUR LA PROTECTION DES DONNÉES

# OSINT et conformité au RGPD



# Table des matières

# Introduction

L'OSINT, ou « Open Source Intelligence » désigne une pratique consistant à collecter et analyser des informations librement accessibles afin de produire des renseignements exploitables. Ces données proviennent d'une multitude de sources, souvent sur Internet, des bases de données publiques, ou encore des réseaux sociaux.

À l'origine, l'OSINT et les technologies qui y sont associées, telles que les services web ou les réseaux cachés du darknet, ont été développées par les agences de renseignement pour répondre à des besoins spécifiques : surveillance, sécurité nationale, et lutte contre le terrorisme[1]. Cependant, son utilisation s'est largement étendue au-delà du cadre gouvernemental. Aujourd'hui, l'OSINT est devenue un outil incontournable pour de nombreux acteurs du secteur privé, couvrant une multitude de domaines. En journalisme, elle est fréquemment employée pour l'investigation et la vérification des faits. Dans le monde de l'entreprise, elle sert à atteindre divers objectifs stratégiques, allant de la gestion des ressources humaines à l'intelligence économique et à la veille concurrentielle. Par ailleurs, la cybersécurité s'appuie largement sur les techniques d'OSINT pour identifier et prévenir les menaces, renforçant ainsi la protection des systèmes et des données sensibles.

Cependant, la pratique de l'OSINT se répand dans des domaines tels que le marketing, les ressources humaines ou encore la gestion des risques des problématiques liées à la protection des données personnelles. Avec la généralisation de son utilisation, notamment dans le secteur privé, le respect des cadres réglementaires comme le RGPD (Règlement général sur la protection des données) devient un impératif incontournable pour concilier efficacité et conformité juridique.

Dans cet article, nous nous concentrerons exclusivement sur les usages civils et privés de l'OSINT, en excluant ses applications dans les domaines de la sécurité nationale et du renseignement d'État. Nous explorerons les applications pratiques de l'OSINT dans le contexte de l'entreprise. Ensuite, nous aborderons les défis juridiques et de conformité au RGPD.

# OSINT, l'intelligence open source déchiffrée

## *Clarification des concepts*

La distinction entre OSINT, Open Source et Open Data est nécessaire.

Tout d'abord, il convient de préciser les notions sous-jacentes à l'acronyme. L'OSINT consiste à collecter des données provenant de diverses sources, telles que les médias, publications, réseaux sociaux, articles de presse, sous différentes formes (textes, images, audio, métadonnées, données de géolocalisation), afin de déduire des informations pertinentes sur un sujet spécifique en fonction d'un objectif initialement fixé.

D'une part, 'OS' signifie 'Open Source', mais il ne fait en aucun cas référence aux ressources telles que les logiciels, bibliothèques ou algorithmes 'libres' couramment utilisés en intelligence artificielle. L'Open Source désigne un type de logiciel dont le code source est rendu public et peut être librement utilisé, modifié et redistribué par quiconque sur des plateformes telles que « GitHub » ou « Hugging Face» . L'objectif de ces plateformes de collaboration et de partage est de permettre à n'importe qui d'examiner, améliorer et partager le code, afin de favoriser l'innovation. Toutefois, il n'est pas exclu dans le cadre d'une investigation numérique, qu'il soit possible de faire usage de ces outils et ressources pour atteindre ses objectifs[2]. Il convient tout de même de préciser qu'il s'agit ici de deux concepts distincts.

D'autre part, il est également pertinent de souligner que la notion de « données publiquement accessibles » induite par la pratique d'OSINT est à distinguer d'autres notions. En effet, les termes « open data » ou « données ouvertes » sont spécifiques à la mise à disposition et la publication de données ouvertes par les entités publiques. Ici encore, tout en soulignant qu'il s'agit de sources différentes, il n'est pas exclu que ces données soient utilisées et exploitées dans un contexte d'investigation numérique. En effet, il est admis que les sources et les données utilisées en OSINT puissent être aussi diverses et variées que les contextes dans lesquels elles sont appliquées et les objectifs initialement fixés.

[1] Open-Source Intelligence Handbook, 2001, NATO

[2] Open-Source Intelligence Tool and Resources Handbook, 2020, Aleksandra Bielska

## *La méthodologie*

La méthodologie de l'OSINT repose sur un processus structuré qui permet de maximiser l'efficacité de la collecte et de l'analyse d'informations. Ce processus se déroule en plusieurs étapes, chacune jouant un rôle spécifique dans l'obtention de renseignements exploitables.

1. La définition des objectifs consiste à préciser ce que l'on cherche à découvrir, en alignant la collecte d'informations avec les besoins spécifiques.
2. La recherche des sources implique l'identification et la sélection des sources pertinentes, telles que les bases de données publiques, les réseaux sociaux, et les sites web.
3. La collecte des données suit, durant laquelle les informations sont extraites et rassemblées de manière systématique.
4. L'analyse des données permet de traiter et d'interpréter les informations recueillies pour en extraire des informations utiles et identifier les éventuelles inconsistances ou incohérences causées par des campagnes de désinformation ou deepfake.
5. La présentation des résultats consiste à organiser et à communiquer les conclusions de manière claire et exploitable pour la prise de décision et/ou initier les actions nécessaires.

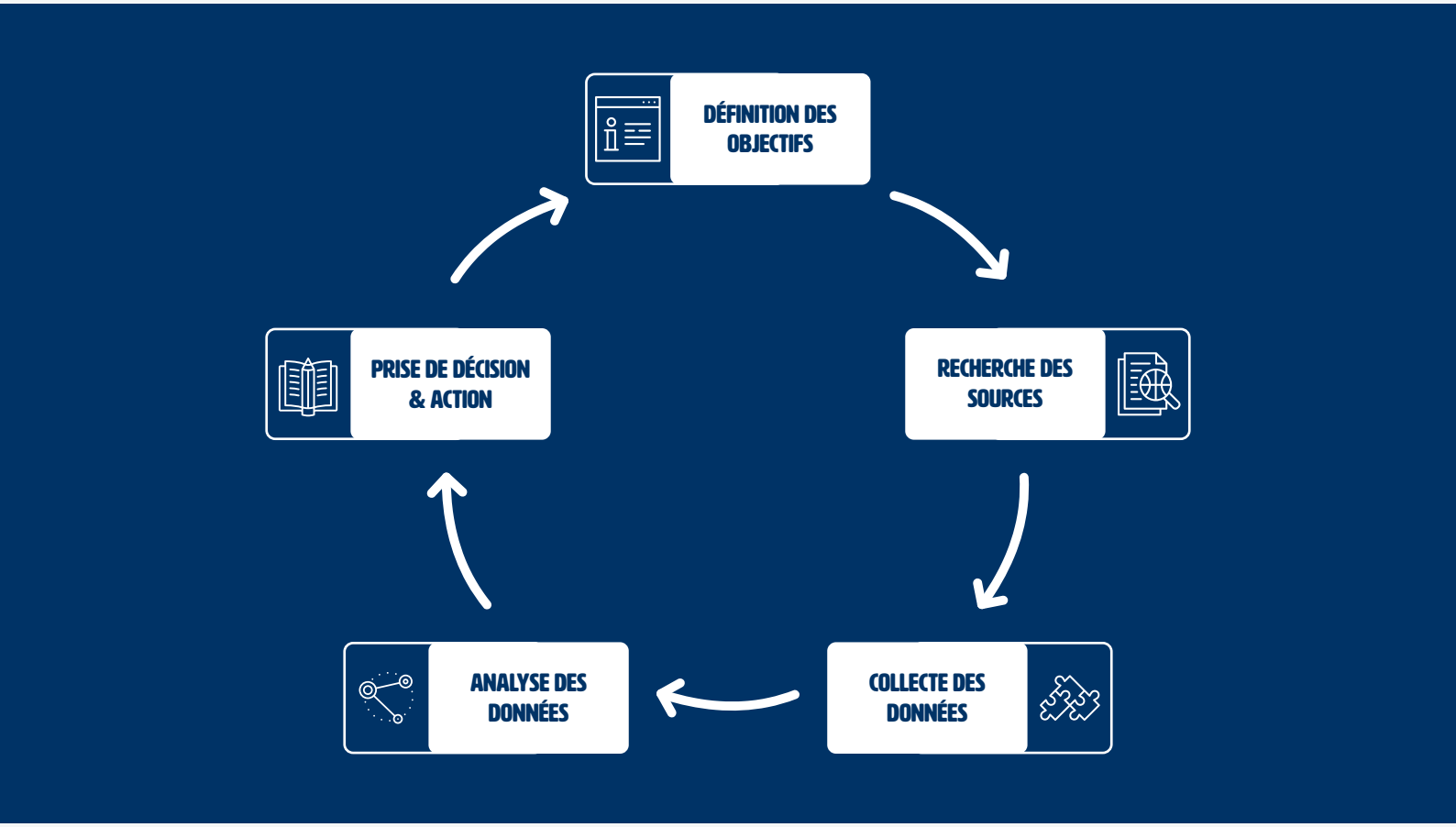



Illustration des étapes du cycle méthodologique de l'OSINT [3]

[3] https://cheapsslsecurity.com/blog/5-best-osint-tools-for-research-penetration-testing/

## L'OSINT et la protection des données à caractère personnel

L'OSINT implique la collecte et l'analyse d'informations publiques, ce qui soulève des enjeux importants en matière de protection des données personnelles et représente un risque de violation de la vie privée. À vrai dire, le fait qu'une donnée soit publiquement accessible ne signifie pas nécessairement qu'elle peut être librement collectée, exploitée et traitée par la suite.

Une telle démarche pourrait contrevenir aux principes énoncés par le RGPD, notamment celui de minimisation, de limitation et même de licéité. Le règlement pose certaines limites et impose aux responsables du traitement de garantir une proportionnalité dans la collecte des données, notamment uniquement les données nécessaires à l'accomplissement d'une finalité clairement définie et légitime permettant de garantir le respect de la vie privée.

Chaque traitement de données doit être justifié par une finalité précise, qu'il s'agisse de la prévention de menaces ou de l'analyse concurrentielle. Il n'existe malheureusement pas de règle universelle pour garantir la conformité aux réglementations en vigueur, telles que le RGPD, car celle-ci est étroitement liée au contexte d'utilisation. Nous allons, dans ce qui suit, présenter quelques cas d'usage illustrant cette nécessaire contextualisation. Nous aborderons la pratique de l'OSINT dans le marketing, le recrutement et la gestion des risques.

## L'OSINT dans l'entreprise : les cas d'usage

### *Le marketing et analyse de marché*

L'OSINT offre aux entreprises un moyen d'analyser les préférences des consommateurs et les tendances du marché à travers la surveillance des réseaux sociaux, l'analyse des avis en ligne et la veille concurrentielle. En effet, avec l'utilisation des réseaux sociaux comme X, Facebook ou Instagram, les entreprises peuvent évaluer l'intérêt des consommateurs et suivre les hashtags populaires, tout en surveillant les campagnes des concurrents. La combinaison avec d'autres sources d'information en ligne, notamment les avis sur les sites de commerce électronique (Amazon, Yelp) permet alors aux entreprises de mieux comprendre les attentes et les perceptions des clients. Cette approche inclut également l'identification d'influenceurs ayant un impact sur le marché cible.

Le contexte métier suggère un intérêt légitime dans la collecte de toutes ces données ; toutefois, la licéité de ce traitement dépendrait d'une évaluation de l'impact sur la vie privée des personnes concernées.

### *Le recrutement*

Le domaine des ressources humaines est particulièrement représentatif de l'usage courant de ces pratiques. Une étude allemande, menée par Deloitte en 2012, indiquait que 13 %[4] des entreprises cotées sur les indices DAX et MDAX avaient déjà recours à des vérifications en ligne et à des contrôles d'antécédents pour leurs candidats à l'embauche. En France, un sondage moins formel mené par le journal « Les Échos » annonce que jusqu'à 85 % des responsables RH affirmaient « googliser » un candidat pour vérifier les informations figurant sur son CV[5]. Cependant, certaines recherches qui peuvent s'apparenter à des investigations sur les candidatures risquent d'être disproportionnées et de porter atteinte au droit au respect de la vie privée.

[4] Studie-Risikominimierung-bei-der-Personalauswahl.pdf (deloitte.com)

[5] https://www.lesechos.fr/idees-debats/leadership-management/85-des-recruteurs-font-des-recherches-en-ligne-sur-les-candidats-1246103

### *La gestion des risques*

Dans un contexte général de gestion des risques cyber, les entreprises surveillent et analysent les données publiques les concernant afin d'améliorer leur évaluation des risques et de renforcer leur posture de sécurité, notamment pour protéger leurs secrets d'affaires. L'utilisation d'outils de surveillance proactive d'Internet, tels qu'« IntelligenceX » permet d'évaluer les vulnérabilités des entreprises en se basant sur les données et informations exposées en ligne.

Cette approche permet de détecter rapidement d'éventuelles fuites de données sensibles et contribuer à la conformité réglementaire, notamment en ce qui concerne l'article 33 du RGPD. De fait, cette disposition exige que le responsable du traitement identifie les fuites de données et, dans certains cas de divulgation directe en ligne, notifie immédiatement l'autorité de protection des données dans un délai de 72 heures[6].

Par ailleurs, des évaluations techniques des surfaces exposées et vulnérables peuvent être réalisées dans le cadre de prestations de services, notamment la sous-traitance. Les entreprises spécialisées en cybersécurité sont souvent engagées par des sociétés qui ne possèdent pas les compétences internes nécessaires pour assurer un niveau approprié de cybersécurité. Cela inclut des prestations de sécurité offensive telles que les tests d'intrusion (Pentest), les campagnes de phishing pédagogique, ou les exercices de Red Team[7].

## Exigences générales du RGPD en matière d'OSINT

### *La transparence et conformité des traitements de données est requise*

Les entités qui utilisent l'OSINT pour collecter et traiter des données doivent inclure ces activités dans leur registre des activités de traitement conformément à l'article 30 du RGPD. Elles sont ainsi soumises aux mêmes obligations qu'un responsable du traitement. Cela signifie qu'elles doivent se conformer aux exigences du règlement en veillant à respecter les principes fondamentaux de transparence et de minimisation des données. En pratique, cela implique d'informer clairement les individus sur la façon dont leurs données sont collectées et utilisées (Art. 13), tout en limitant la collecte aux informations strictement nécessaires (Art 5-1-c) et en mettant en place des mesures de sécurité adaptées pour protéger ces données (Art 32). De plus, la durée de conservation des informations doit être précisément délimitée pour éviter toute rétention excessive ou injustifiée des données personnelles (Art 5-1-e).

### *Les limites*

La pratique de l'OSINT peut être exercée en conformité avec le RGPD sous certaines conditions précises, telles que l'existence d'une base légale appropriée, la minimisation des données collectées, et la limitation de la durée de rétention. Cependant, elle est également soumise à des restrictions strictes en ce qui concerne l'acquisition, l'achat ou la réutilisation de données exfiltrées et publiées illégalement. En effet, certaines données sensibles, bien que non accessibles via le web traditionnel, circulent sur des réseaux parallèles comme le darknet. Leur exploitation constitue non seulement une violation des principes éthiques et une non-conformité au RGPD, mais également une infraction pénal[8]. Ces informations, souvent obtenues illégalement à la suite de piratages informatiques, sont ensuite mises en vente par des cybercriminels.

Il est vrai que cette approche d'investigation numérique offre un accès à des ressources exploitables pour des activités illicites. En fait, des acteurs malveillants peuvent utiliser ces approches pour notamment collecter des informations en vue de planifier des cyberattaques, lancer des campagnes de phishing sophistiquées, réaliser des « fraudes au président », orchestrer des extorsions ou autres activités manifestement illégales.

---

[6] Lignes directrices 01/2021 EDPB Exemples concernant la notification de violations de données à caractère personnel

[7] Article 6 du RGPD : Selon l'article 6 du RGPD, la base légale de l'exécution contractuelle peut être envisagée dans un contexte de sous-traitance à des sociétés spécialisées lors de l'exécution d'une prestation de service de cybersécurité. Néanmoins, le responsable du traitement doit évaluer la base légale adéquate à la lumière du considérant 49, notamment dans le contexte de l'identification des risques cyber après avoir réalisé un exercice de mise en balance approprié avant de confier ce type de traitement à ces sociétés spécialisées.

[8] Article 509-4 du code pénal [LU]

## Conclusion et recommandations

Comme pour toute activité au sein de l'entreprise, il est crucial de privilégier une gouvernance adéquate. Les activités métier et de support doivent être encadrées par les politiques et procédures internes, ainsi que par les contrats de sous-traitance lorsqu'elles sont externalisées. Les procédures relatives à la recherche, la récupération et l'exploitation des données collectées dans le cadre des missions doivent être définies. En plus, des mesures de précaution doivent être mises en place pour garantir la confidentialité des vulnérabilités identifiées et pour limiter la durée de stockage des informations, afin de protéger efficacement les données sensibles[9].

Par l'intégration des standards (ISO/IEC 27001 pour la gestion de la sécurité de l'information) dans les politiques et procédures internes, l'entreprise pourra non seulement améliorer la gestion de ses données, mais aussi renforcer une conformité aux exigences réglementaires en exerçant une veille continue sur les standards et les ressources publiés par les institutions spécialisées ci-dessous (la liste est non exhaustive).

- NIST (National Institute of Standards and Technology) pour des lignes directrices complètes en cybersécurité ;
- CIS (Center for Internet Security) pour des pratiques de sécurité éprouvées et des contrôles de sécurité recommandés ;
- ENISA (Agence Européenne pour la Cybersécurité) pour des rapports de recherche, des guides et bonnes pratiques, des normes et Framework, ainsi que des alertes et avertissements concernant les menaces émergentes et les vulnérabilités ;
- EDPB (European Data Protection Board) pour les lignes directrices publiées concernant la protection des données et les guidances des autorités de supervision.

Bien que la veille technologique soit aujourd'hui largement intégrée dans les stratégies des organisations, elle doit être accompagnée d'une veille réglementaire et d'une observation des bonnes pratiques. Ces dimensions permettent de bâtir une capacité à anticiper et répondre efficacement aux enjeux de conformité dans un environnement numérique en constante évolution, où les avancées technologiques s'accélèrent.

En somme, les organisations, qu'elles soient publiques ou privées, doivent adopter une posture proactive et vigilante face aux risques inhérents à l'exposition croissante d'informations en ligne. Une attention renforcée et une gestion prudente de ces enjeux ne protègent pas uniquement la vie privée des individus, mais elles constituent également un pilier fondamental de la sécurité globale des organisations, garantissant leur résilience et leur pérennité dans un monde numérique complexe et interconnecté.

[9] Article 25 du RGPD