AYI NEDJIMI Consultants

Classification des données : méthodes et outils pratiques

Catégorie : Conformité Lecture : 9 min Publié le : 12/03/2026 Auteur : Ayi NEDJIMI

Classifiez vos données avec une méthodologie éprouvée. Niveaux de sensibilité, outils de data discovery, étiquetage automatisé et déploiement.

Résumé exécutif

La classification des données est le prérequis fondamental de toute stratégie de protection de l'information efficace, permettant d'appliquer des mesures de sécurité proportionnées à la sensibilité réelle des données traitées plutôt que d'imposer un niveau de protection uniforme et souvent inadapté à l'ensemble du patrimoine informationnel. Ce guide détaille les méthodologies de classification reconnues, les niveaux de sensibilité à définir en fonction du contexte organisationnel et réglementaire, les outils techniques de découverte et d'étiquetage automatisé des données disponibles sur le marché, et les bonnes pratiques de déploiement d'un programme de classification pérenne couvrant l'ensemble des référentiels de données structurées et non structurées de l'organisation dans ses environnements on-premise et cloud, en s'appuyant sur les recommandations normatives ISO 27001 et les exigences réglementaires RGPD et NIS 2 en matière de protection proportionnée de l'information.

La classification des données est paradoxalement l'une des mesures de sécurité les plus fondamentales et les plus négligées dans les organisations françaises et européennes confrontées à une croissance exponentielle du volume de données numériques qu'elles produisent, traitent et stockent quotidiennement. Sans *classification* formelle et opérationnelle, il est impossible de déterminer quelles données méritent un chiffrement systématique, quelles données peuvent être stockées dans le cloud public, quelles données nécessitent des contrôles d'accès renforcés et quelles données doivent faire l'objet d'une surveillance spécifique en matière de prévention des fuites. Le résultat est un paradoxe coûteux : soit l'organisation applique des mesures de sécurité maximales à toutes les données, ce qui génère des coûts prohibitifs et des frictions opérationnelles insoutenables, soit elle applique des mesures minimales uniformes, exposant ses données les plus sensibles à des risques inacceptables. La **norme ISO 27001** exige explicitement la classification de l'information via le contrôle A.5.12, la directive NIS 2 impose des mesures proportionnées au niveau de risque, et le **RGPD** requiert une protection renforcée des catégories particulières de données personnelles définies à l'article 9. La mise en place d'un programme de classification structuré n'est donc pas seulement une bonne pratique de sécurité mais une obligation réglementaire transverse dont la mise en œuvre effective conditionne la conformité simultanée à plusieurs cadres normatifs et législatifs.

Pourquoi la classification des données est-elle indispensable ?

La classification des données répond à quatre objectifs stratégiques complémentaires pour toute organisation soucieuse de protéger efficacement son patrimoine informationnel. Le premier objectif est la **proportionnalité des mesures de sécurité** : en connaissant le niveau de sensibilité de chaque catégorie de données, l'organisation peut appliquer des contrôles proportionnés évitant à la fois la sous-protection des données critiques et la surprotection coûteuse des données non sensibles. Le deuxième objectif est la **conformité réglementaire** : le RGPD, NIS 2, DORA et les réglementations sectorielles exigent tous des mesures adaptées au niveau de risque et de sensibilité des données traitées.

Le troisième objectif est la **gestion du cycle de vie des données** : la classification détermine les durées de conservation, les conditions de stockage, les modalités de transfert et les procédures de destruction des données à chaque étape de leur cycle de vie. Le quatrième objectif est la **facilitation de la gestion des incidents** : en cas de fuite de données, la classification permet d'évaluer rapidement la gravité de l'incident et de déterminer les obligations de notification applicables. Ces objectifs s'articulent avec la protection technique des données RGPD et le cadre NIS 2.

Pourriez-vous identifier en moins de trente minutes l'ensemble des emplacements où sont stockées vos données les plus sensibles, y compris les copies sur les postes de travail, les services cloud non référencés et les boîtes email des collaborateurs ?

Comment définir les niveaux de classification adaptés ?

La définition des niveaux de classification doit être adaptée au contexte spécifique de l'organisation tout en restant suffisamment simple pour être comprise et appliquée par l'ensemble des collaborateurs. Un schéma de classification trop complexe avec de nombreux niveaux sera systématiquement mal appliqué sur le terrain. La plupart des organisations adoptent un schéma à **quatre niveaux** : public (données destinées à être diffusées sans restriction), interne (données réservées aux collaborateurs de l'organisation), confidentiel (données dont la divulgation pourrait causer un préjudice significatif) et secret (données dont la divulgation pourrait causer un préjudice grave).

Chaque niveau doit être accompagné d'une **définition claire** illustrée par des exemples concrets tirés du contexte de l'organisation, d'une *matrice de mesures de sécurité* précisant les contrôles applicables (chiffrement, contrôle d'accès, stockage autorisé, transfert autorisé, impression autorisée, durée de conservation) et d'un processus de classification et de déclassification documenté. La responsabilité de la classification initiale incombe au **propriétaire des données** (data owner), typiquement le responsable métier qui crée ou acquiert les données, tandis que le RSSI définit les mesures de sécurité associées à chaque niveau et le DPO assure la cohérence avec les exigences du RGPD pour les données personnelles.

Mon avis : La classification des données est un sujet où la perfection est l'ennemie du bien. J'ai vu des organisations passer deux ans à concevoir un schéma de classification élaboré avec sept niveaux et des matrices de contrôle de cinquante pages, pour finalement ne jamais réussir à le

déployer parce que trop complexe pour les utilisateurs. Quatre niveaux maximum, des définitions en langage clair avec des exemples concrets, et un outil de marquage intégré dans les applications bureautiques : voilà la recette qui fonctionne sur le terrain.

Quels outils techniques pour la classification des données ?

Les outils de classification des données se répartissent en deux catégories complémentaires : les outils de **data discovery** qui identifient et inventorient automatiquement les données sensibles dans les systèmes d'information, et les outils d'**étiquetage** (data labeling) qui permettent aux utilisateurs et aux processus automatisés de marquer les données selon le schéma de classification défini. Les plateformes leaders du marché incluent **Microsoft Purview Information Protection** (anciennement Azure Information Protection) nativement intégré à l'écosystème Microsoft 365, **Varonis** spécialisé dans la découverte et la protection des données non structurées, et **Forcepoint DLP** combinant classification et prévention des fuites de données.

Les outils de data discovery utilisent des techniques d'analyse de contenu (expressions régulières, empreintes numériques, apprentissage automatique) pour identifier les données sensibles telles que les numéros de carte bancaire, les numéros de sécurité sociale, les données médicales et les informations confidentielles d'entreprise. L'intelligence artificielle améliore progressivement la précision de la détection en réduisant les faux positifs et en identifiant les données sensibles contextuelles qui échappent aux règles prédéfinies. Ces outils doivent être intégrés dans l'architecture du SOC et du système de log management pour corréler les événements de sécurité avec le niveau de sensibilité des données concernées.

Niveau de classification	Exemples de données	Chiffrement requis	Contrôle d'accès	Stockage autorisé
Public (C0)	Communiqués de presse, plaquettes commerciales	Non requis	Aucune restriction	Tous supports
Interne (C1)	Procédures internes, organigramme, annuaire	En transit recommandé	Authentification requise	Environnements internes et cloud approuvé
Confidentiel (C2)	Données clients, données financières, code source	Au repos et en transit	Besoin d'en connaître vérifié	Environnements sécurisés approuvés
Secret (C3)	Plans stratégiques, données M&A, secrets industriels	Chiffrement renforcé systématique	Liste nominative validée par direction	Environnements hautement sécurisés uniquement

La fuite de données massive subie par Marriott International entre 2014 et 2018, exposant les données personnelles de 500 millions de clients de la chaîne Starwood, illustre les conséquences catastrophiques de l'absence de classification et de protection différenciée des données. Les données de passeports, de cartes bancaires et de programmes de fidélité étaient stockées sans classification formelle ni mesures de protection proportionnées à leur sensibilité. Une

classification rigoureuse aurait imposé un chiffrement systématique des données de passeports et de cartes bancaires, des contrôles d'accès renforcés et une surveillance spécifique qui auraient considérablement limité le volume et l'impact de la fuite, en lien avec la gestion des vulnérabilités.

Comment déployer un programme de classification à grande échelle ?

Le déploiement d'un programme de classification des données à l'échelle de l'organisation est un projet de transformation qui nécessite une approche progressive et pragmatique pour éviter l'échec classique du projet trop ambitieux qui s'enlise sous son propre poids. La première phase, d'une durée de deux à trois mois, consiste à **définir le cadre** : schéma de classification, politique associée, rôles et responsabilités, et sélection des outils. La deuxième phase déploie le programme sur un **périmètre pilote** ciblant une ou deux directions métier volontaires pour valider le schéma et les outils en conditions réelles.

La troisième phase étend progressivement le programme à l'ensemble de l'organisation par vagues successives, chaque vague bénéficiant des retours d'expérience de la précédente. La formation et la sensibilisation des utilisateurs sont critiques à chaque vague : les collaborateurs doivent comprendre pourquoi la classification est nécessaire, comment classifier leurs données au quotidien et quelles sont les conséquences d'une classification erronée ou absente. L'automatisation progressive via les outils de data discovery réduit la dépendance à la classification manuelle et améliore la couverture du programme, conformément aux recommandations de la CNIL en matière d'accountability et aux standards de l'ENISA sur la protection des données.

Faut-il classifier les données dans le cloud différemment ?

La classification des données dans les environnements cloud ne diffère pas fondamentalement dans ses principes mais nécessite des adaptations spécifiques dans sa mise en œuvre technique. Les données hébergées dans le cloud public sont soumises aux mêmes exigences de classification que les données on-premise, mais les mécanismes de contrôle diffèrent : le chiffrement doit être systématique avec une gestion des clés maîtrisée par l'organisation (BYOK ou HYOK), les contrôles d'accès doivent être alignés avec les politiques IAM du cloud provider, et la localisation géographique des données doit être vérifiée pour les données sensibles soumises à des restrictions de transfert international.

Les architectures multi-cloud complexifient la gestion de la classification car chaque fournisseur propose ses propres outils et mécanismes d'étiquetage qui ne sont pas nécessairement interopérables. L'adoption d'une **solution de classification centralisée** capable de couvrir les environnements on-premise, les différents clouds publics et les applications SaaS de l'organisation est recommandée pour maintenir la cohérence et la visibilité. La politique de classification doit explicitement définir les niveaux de données autorisés dans chaque type

d'environnement cloud, avec des contrôles compensatoires pour les données sensibles dont l'hébergement cloud est nécessaire pour des raisons opérationnelles mais qui nécessitent une protection renforcée.

Comment mesurer l'efficacité du programme de classification ?

La mesure de l'efficacité du programme de classification des données repose sur un ensemble d'indicateurs quantitatifs et qualitatifs permettant de suivre la progression du déploiement et d'identifier les domaines nécessitant des actions correctives ou de renforcement. Les indicateurs quantitatifs clés incluent le taux de couverture des données classifiées par rapport au volume total estimé de données de l'organisation, le taux de données correctement classifiées vérifié par des audits d'échantillonnage périodiques, le nombre de données sensibles découvertes hors périmètre classifié par les outils de data discovery, et le délai moyen de classification des nouvelles données créées ou acquises par l'organisation.

Les indicateurs qualitatifs couvrent le niveau de compréhension et d'adhésion des collaborateurs au schéma de classification mesuré par des enquêtes et des tests pratiques, la cohérence de la classification entre les différentes directions de l'organisation vérifiée par des audits croisés, et l'efficacité des mesures de protection associées à chaque niveau de classification évaluée par des tests techniques ciblés. Ces indicateurs alimentent le tableau de bord du RSSI et sont présentés régulièrement au comité de pilotage sécurité pour démontrer la progression du programme et justifier les investissements nécessaires à son maintien et à son amélioration continue dans le temps.

Sources et références : CNIL · ANSSI

Quelles erreurs classiques éviter dans la classification ?

Les retours d'expérience terrain permettent d'identifier plusieurs erreurs récurrentes qui compromettent l'efficacité des programmes de classification des données et qu'il convient d'anticiper et d'éviter. La première erreur est la surclassification systématique qui conduit les collaborateurs à classifier toutes les données au niveau le plus élevé par précaution, rendant la classification inopérante car elle ne distingue plus les données réellement sensibles des données courantes. La deuxième erreur est l'absence de processus de déclassification qui empêche de réduire le niveau de protection des données dont la sensibilité diminue dans le temps.

La troisième erreur est la déconnexion entre la classification et les mesures de protection effectives : classifier les données n'a de valeur que si des contrôles techniques sont effectivement appliqués en fonction du niveau de classification, ce qui nécessite une intégration avec les outils de DLP, de chiffrement et de contrôle d'accès. La quatrième erreur est le manque de formation et de sensibilisation des propriétaires de données qui doivent comprendre les critères de classification et savoir les appliquer concrètement dans leur contexte métier quotidien. La cinquième erreur est l'absence de gouvernance du programme avec un comité de classification qui arbitre les cas litigieux et maintient la cohérence du schéma dans le temps face aux évolutions organisationnelles et réglementaires.

À retenir : La classification des données est le fondement de toute stratégie de protection de l'information. Adoptez un schéma simple à quatre niveaux maximum, déployez-le progressivement par vagues avec un pilote initial, automatisez la découverte et l'étiquetage avec des outils adaptés et formez tous les collaborateurs à classifier leurs données au quotidien. La classification n'est pas un projet ponctuel mais un processus continu intégré dans la culture de l'organisation.

Ayi NEDJIMI Consultants — Expert cybersécurité offensive & intelligence artificielle

ayinedjimi-consultants.fr · ayi@ayinedjimi-consultants.fr