

MICROPOLÉ

Les clés de succès du Data Cleaning

LIVRE BLANC

2024

edito

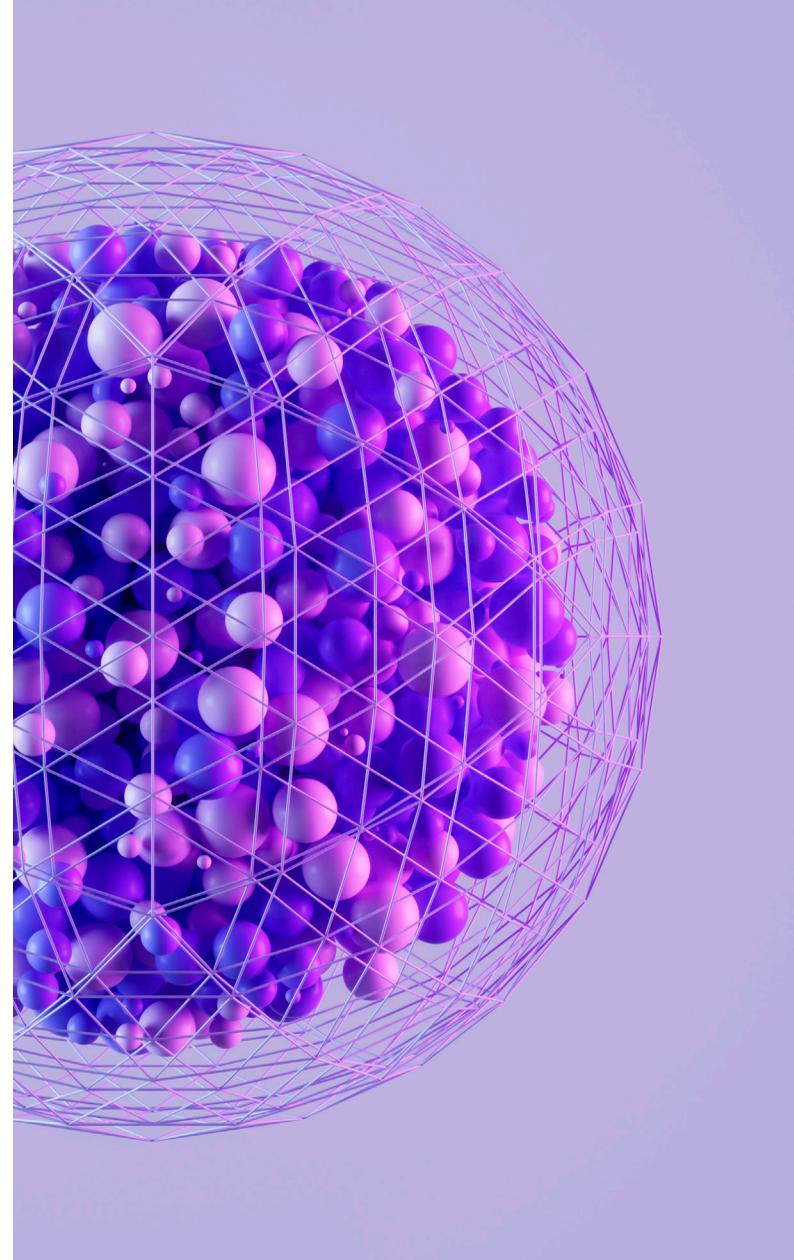


ELIOTT MOURIER

*Partner Data Compliance & Privacy,
Micropole*



Maîtriser ses données, c'est aussi savoir quand, où et comment les nettoyer lorsqu'elles atteignent la fin de leur cycle de vie. Si les contraintes réglementaires en la matière sont de plus en plus fortes (notamment en vue de minimiser les risques de fuites de données), le Data Cleaning est également un enjeu majeur pour s'assurer que votre patrimoine de données conserve sa pertinence et sa valeur dans le temps.



sommaire

- 3 raisons qui rendent le Data Cleaning incontournable 04
- Le cycle de vie de la donnée 09
- 5 conseils pour définir les règles de conservation des données 13
- Réussir son projet de Data Cleaning 18



3 raisons qui rendent le Data Cleaning incontournable



Alors que la donnée est omniprésente autant dans le business que dans notre vie privée, les sujets liés à la Data Retention (conservation des données) et au Data Cleaning (nettoyage des données) deviennent de véritables enjeux et priorités pour les entreprises. Ce constat est renforcé par un triple contexte :

L'augmentation exponentielle des données

Partout, le volume de données explose. Les propriétaires de ces données se retrouvent confrontés à un défi perpétuel concernant leur gouvernance.

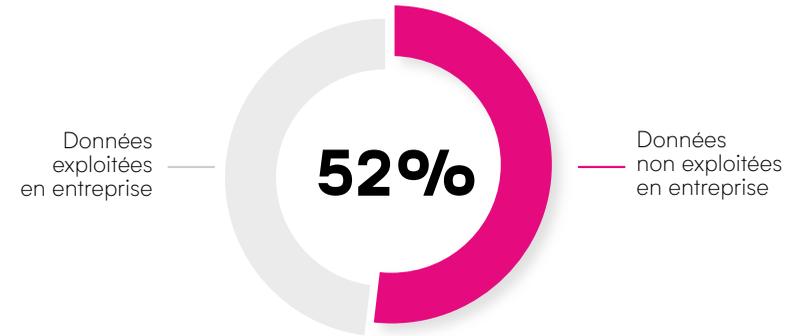
La tendance est plus que jamais à la hausse. En 2021, l'institut de statistiques allemand Statista¹ estimait que le volume de données générées devrait dépasser les 180 zettaoctets d'ici 2025. En 2030, selon un rapport édité par Huawei, on pourrait compter le volume mondial de **données générées en yottabits²** (10^{24} bits).

Cette prise de conscience de l'explosion du volume de données s'accompagne tout naturellement d'une recherche de solution.

Une pression réglementaire grandissante

Les données sont de plus en plus nombreuses, mais également de plus en plus **encadrées par la loi**. Diverses réglementations européennes et nationales obligent les organisations à **maîtriser le cycle de vie** de leurs données.

Il peut être compliqué de s'y retrouver parmi les obligations de destruction, d'archivage, d'anonymisation mais aussi de durées de rétention légales qui sont soumises à sanction.



La nécessité d'une sobriété énergétique

L'impact environnemental du stockage des données constitue un **enjeu sociétal majeur**. On estime l'empreinte carbone du stockage d'un 1 To de données sur des serveurs en France à 1,760 tonne d'équivalent CO₂ par an.

Une étude de Veritas³ parue en 2020 estime également qu'en moyenne, 52% des données d'une entreprise sont inutilisées, entraînant un rejet de 6,4 millions de tonnes de CO₂.

Data Cleaning & RSE

La gestion des données ne se limite pas à l'aspect économique, elle doit également prendre en compte des aspects environnementaux et sociaux. Conserver des données sans plan d'utilisation clair engendre souvent des coûts financiers et environnementaux considérables.

D'après l'Agence de l'Environnement et de la Maîtrise de l'Energie, les centres de données sont à l'origine de **2% des émissions de gaz à effet de serre** mondiales.

Les entreprises doivent élaborer des politiques de stockage et de conservation alignées sur des **principes de durabilité**, pour minimiser leur empreinte carbone et favoriser une utilisation plus efficiente des ressources.

L'adoption de règles de conservation intégrant le **numérique responsable** contribue à un équilibre entre les impératifs économiques et les responsabilités environnementales.

Ce travail peut être valorisé dans le cadre de la mise en conformité à la directive **CSRD**, entrée en vigueur en début d'année, qui impose aux entreprises de mesurer leurs impacts ESG (Environnementaux, Sociaux et de Gouvernance) et de justifier les actions menées pour s'améliorer.



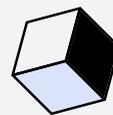
Les solutions apportées par le Data Cleaning



Maîtriser la
croissance du volume
de données



Faciliter la gestion
et la gouvernance
de la donnée

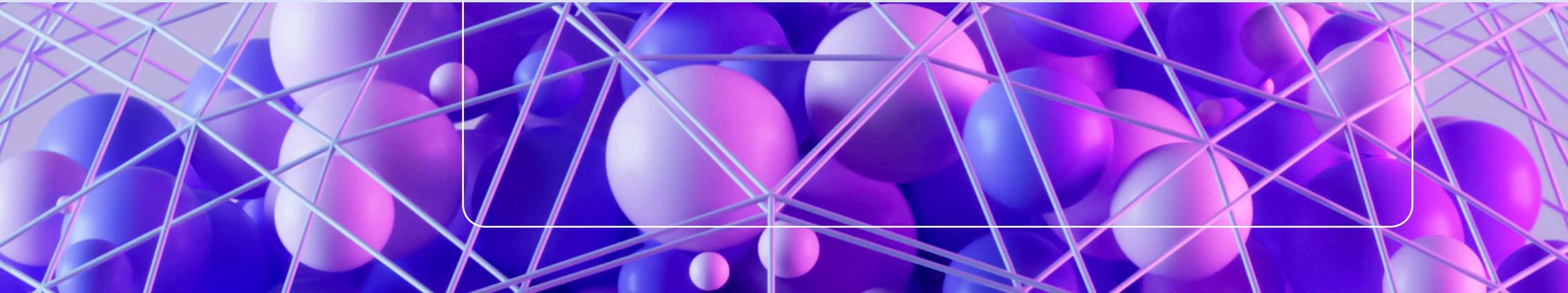


Se conformer aux
réglementations en
vigueur

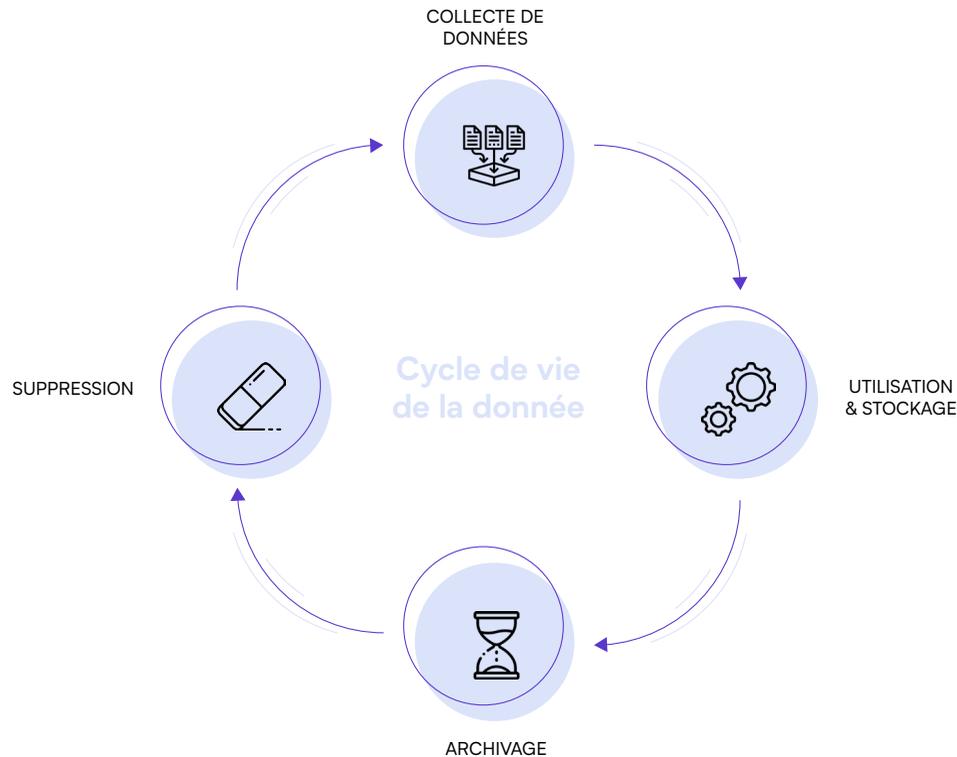


Contribuer à réduire /
rationaliser l'empreinte
carbone des entreprises

Le cycle de vie de la donnée



Bien définir les étapes du cycle et leurs objectifs respectifs permettra de s'assurer qu'à tout moment les données sont traitées et consultées conformément aux engagements de l'entreprise et contraintes légales. Ceci limitera également les risques d'accès inappropriés ou de fuites de données.



Collecte de données

Les données sont collectées selon des processus métiers bien définis et par des systèmes adaptés. Elles doivent toujours être minimisées tout en étant suffisantes pour satisfaire les besoins de l'entreprise.



Utilisation & stockage

Les données sont stockées dans les systèmes et utilisées par les différents processus et systèmes de l'entreprise aussi longtemps que nécessaire.



Archivage

A un moment déterminé, souvent défini par la loi ou des normes, les données devenues « froides » peuvent être archivées ou leur accès peut être restreint.



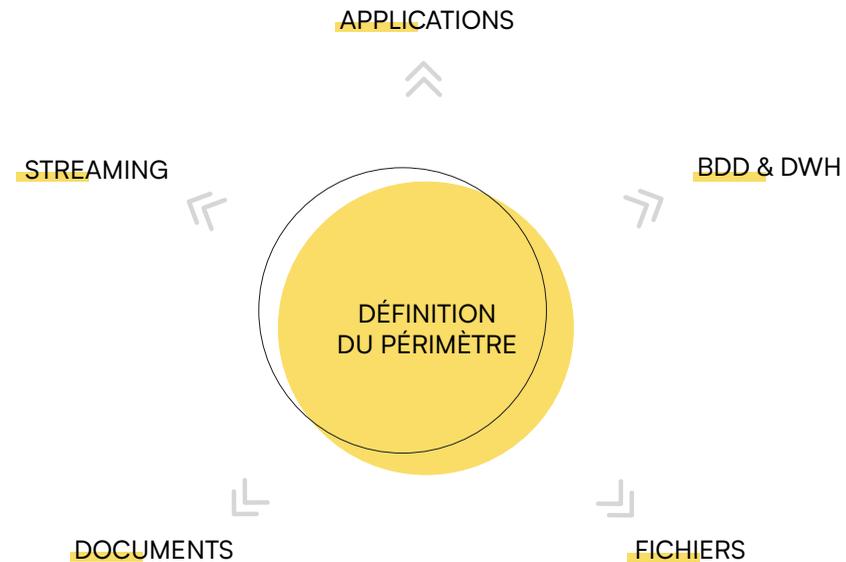
Suppression

A l'issue des délais de conservation légaux ou convenus par l'entreprise, les données sont purgées ou anonymisées.

Analyser le périmètre c'est également considérer toutes les typologies de données qui transitent dans l'ensemble de l'organisation. Elles peuvent être relatives aux :

-  **Opérations**
Financières ou de gestion
-  **Communications**
E-mails, conversations...
-  **Données de référence**
Référentiels, RH...
-  **Analyses & décisions**
Reportings...
-  **Contenus & multimédia**
Logos, vidéos, marketing...
-  **Traces & notifications**
Workflows...

Méthodologie et pragmatisme sont nécessaires pour mener à bien un projet de Data Cleaning. La première étape consiste à définir le périmètre général du projet : en termes de supports concernés et de typologies de données.



USE CASE

Gestion & conservation des données de 75 applications

Réponse aux enjeux de gestion et conservation de la donnée sur un ensemble de 75 applications différentes pour un Groupe de service de paiement aux entreprises et commerçants.

- Pilotage d'un **programme de purge et d'archivage** des données personnelles (75 applications)
- Inventaire des données personnelles dans l'ensemble du SI, **évaluation du niveau de risques** (volumétrie, sensibilité, exposition, obsolescence, etc.)
- Définition d'une **matrice de règles** communes de purge/archivage
- Construction d'une **roadmap pluriannuelle** chiffrée
- Suivi de la mise en oeuvre des **remédiations dans les systèmes** et production de preuves à des fins d'audit



Données
personnelles



Parc
applicatif



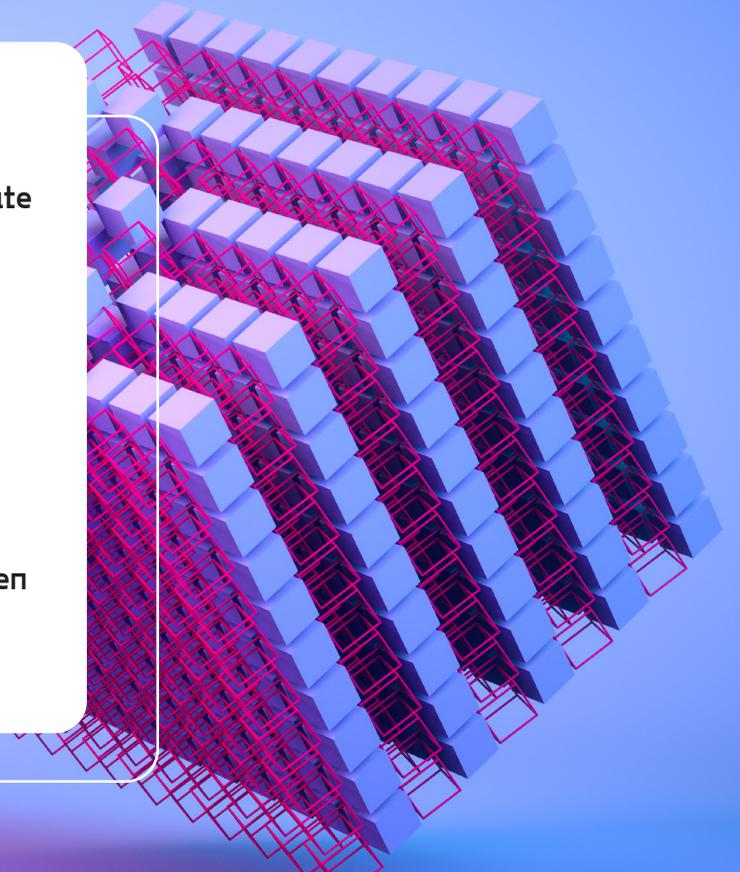
Loi et
conformité

5 conseils pour définir les règles de conservation des données



La durée de conservation des données est généralement définie par la loi. A titre d'illustration, le Code du travail impose que les informations consignées sur le registre unique du personnel soient conservées pendant 5 ans à compter de la date à laquelle le salarié a quitté l'établissement.

Néanmoins, il existe de nombreuses situations pour lesquelles la durée de conservation des données n'est pas explicitement définie par la loi. Dans ces cas de figure, il incombe à l'entreprise de déterminer une période de conservation proportionnée à la finalité du traitement. Les entreprises doivent ainsi construire leur propre référentiel de règles de conservation en tenant compte de l'objectif initial de la collecte de données, de la finalité du traitement, des réglementations en vigueur, mais aussi de leurs spécificités métiers et sectorielles ou encore de leurs contraintes techniques.



1

Respecter les périodes légales de stockage des données

Pour déterminer les périodes de conservation des données actives ou archivées, une attention doit être accordée à la période prescrite par la loi en vigueur. Il est impératif que dans le traitement de données personnelles, les durées de conservation respectent le principe de limitation du stockage défini par la loi.

2

Définir la durée de conservation en adéquation avec le RGPD

En l'absence de durée de conservation prescrite par la loi, les entreprises doivent définir la période en tenant compte de la politique de protection des données, des délais de prescription légaux, des exigences de contrôle et de l'utilisation des données par d'autres entités de l'organisation. Les impératifs professionnels doivent être réévalués sur la base du principe de minimisation des données du RGPD.

3

Identifier les risques de conflits de conservation et anticiper leur résolution

En cas de conflits ou de manque de clarté entre les exigences réglementaires, une évaluation des risques doit être effectuée, particulièrement lorsqu'il est décidé de conserver les données au-delà des délais légaux. La consultation d'experts juridiques spécialisés dans la protection des données peut s'avérer bénéfique pour les entreprises afin d'assurer la conformité des règles de conservation avec les lois en vigueur.

4

Adapter les périodes de conservation en fonction des finalités

La durée de conservation de certaines données peut être davantage liée aux utilisations spécifiques des données qu'à leur existence même, nécessitant une gestion des accès en fonction des finalités (ex.: données liées à la gestion de contrat vs données marketing). La création d'un accès différencié aux données peut ainsi aider à gérer les conflits.

5

Simplifier le processus de conservation des données

Simplicité et pragmatisme doivent être les maîtres-mots dans la définition et l'implémentation des périodes de conservation, pour éviter des règles complexes susceptibles de générer des difficultés et une exécution complexe. Les process mis en place doivent être bien compris et correctement mis en œuvre par l'ensemble des parties prenantes pour en garantir la conformité.

« La définition des règles de conservation des données doit nécessairement impliquer l'ensemble des parties prenantes de l'entreprise et pas uniquement le département juridique ou le DPO. Tenir compte des usages métiers, des standards en matière de gouvernance de la donnée, des normes propres au secteur d'activité ou encore des contraintes techniques est indispensable pour établir des règles cohérentes qui pourront ensuite être concrètement appliquées et déclinées sur l'ensemble du système d'information. »



Elliott MOURIER
*Partner Data Compliance
& Privacy, Micropole*

USE CASE

Pilotage et coordination d'un programme Data Retention international

Nettoyage d'une trentaine de systèmes critiques (CRM, Datawarehouse, Compliance, etc.) de l'ensemble des données clients clos depuis plus de 10 ans dans le domaine bancaire.

- Définition et validation des **règles de conservation** à appliquer en fonction des finalités et des spécificités locales
- Mise en œuvre des **solutions de purge physique**, d'anonymisation ou d'archivage sur bande dans les différents systèmes
- Exécution des **traitements de purge/anonymisation** sur +100.000 clients clos en Suisse, Monaco, Espagne, Singapour & Hong-Kong



Données clients

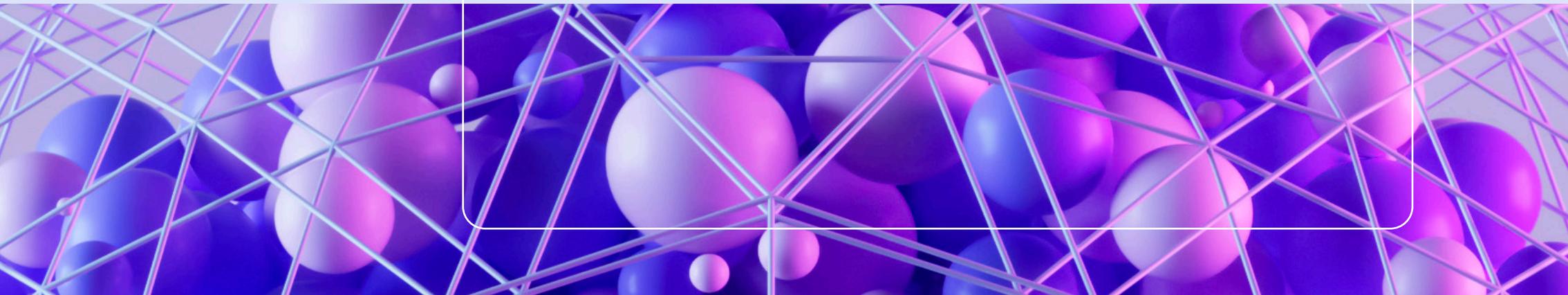


Définition des règles



Purge & anonymisation

Réussir son projet de Data Cleaning



4 étapes pour gérer efficacement le cycle de vie des données

<< Garant de la qualité des données, la Data Governance vise à gérer, exploiter, optimiser, évaluer, protéger, contrôler, et entretenir les données des entreprises.

La fin de vie des données est souvent le parent pauvre de la gouvernance des données alors que les enjeux de compliance sont majeurs pour cette étape. >>



Pascal ANTHOINE
*Data Governance Director
Senior Partner, Micropole*

1

Inventaire & analyse

Recensement des données par types dans les systèmes

2

Définition des règles et délais de rétention

Formalisation des règles de nettoyage et des délais de rétention à appliquer

3

Etude & chiffrage

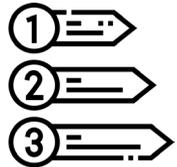
Etude des solutions techniques, chiffrage et planification

4

Mise en oeuvre et rapports

Pilotage et coordination des chantiers de mise en oeuvre, apport d'expertise

Les bénéfices du Data Cleaning



PRIORISATION & QUICK-WINS

Priorisation en fonction des risques, stratégie de lots successifs et de quick-wins pour justifier d'une démarche concrète et efficace.



COLLÉGIALITE & PRAGMATISME

La définition collégiale des règles et des délais de rétention est impérative pour parvenir à allier réglementations et réalités métiers et techniques.



FLEXIBILITÉ TECHNIQUE

Purge physique des données, anonymisation, archivage sécurisé, etc., les options techniques sont nombreuses et peuvent fournir des alternatives satisfaisantes et souvent moins coûteuses.



OPTIMISATION BUDGÉTAIRE

Mutualisation des coûts et des plannings de mise en œuvre des solutions, ainsi que des ressources (devs & tests), recherche du meilleur rapport coût/efficacité/conformité.



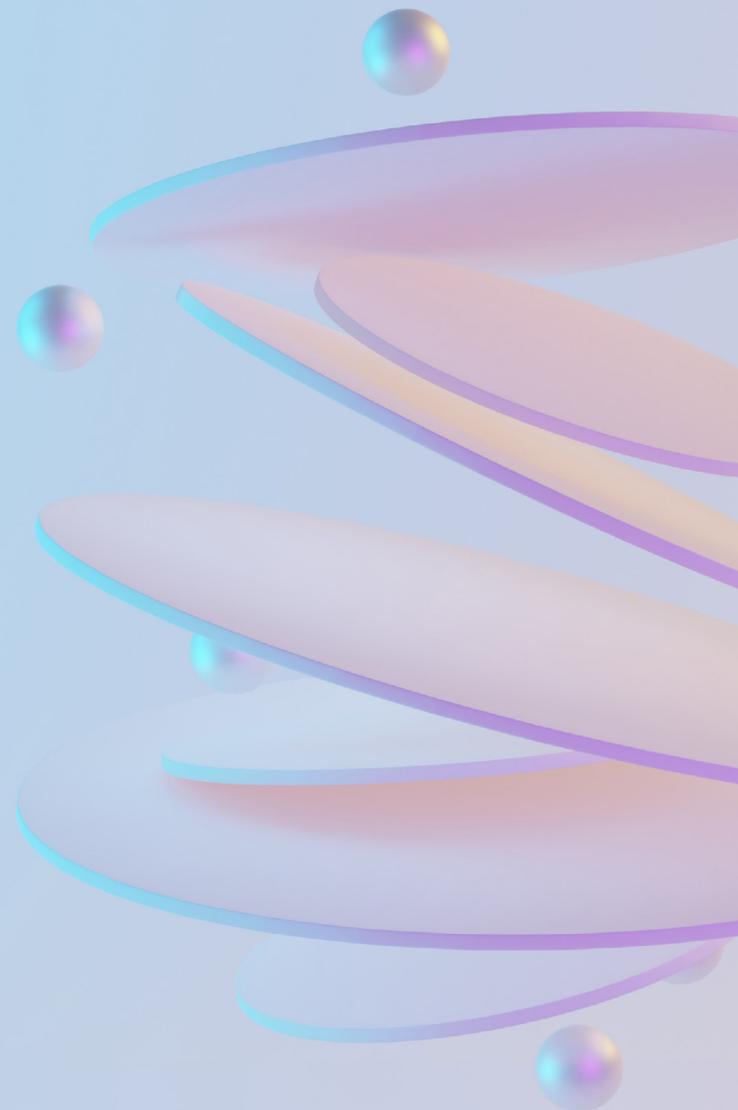
TRACABILITÉ

Les solutions mises en œuvre devront nécessairement permettre une traçabilité fine des actions de remédiation de manière à disposer de preuves en cas d'audit ou de contrôle.

conclusion

Avec la prolifération des réglementations et la pression croissante des sujets environnementaux, le Data Cleaning devient aujourd'hui et plus que jamais un enjeu majeur pour les organisations qui ne peuvent plus simplement accumuler de la donnée sans se soucier de leur sort, mais sont tenues d'en maîtriser le cycle de vie. Si la définition des règles à appliquer n'est pas toujours aisée, leur mise en œuvre concrète sur l'ensemble du système d'information constitue un défi encore plus conséquent.

Seule une approche méthodologique structurée et pragmatique, comme celle que Micropole applique depuis des années maintenant auprès de ses clients, peut aider les entreprises à en venir à bout. Ce qui est certain, c'est que la capacité d'une organisation à construire et mettre en place une démarche de Data Cleaning constituera demain un indicateur majeur de sa capacité à maîtriser son patrimoine data et à en tirer le plein potentiel.





**LE DATA CLEANING EST UN ENJEU
POUR VOTRE ORGANISATION ?**

Contactez nos experts

 info@micropole.com

RÉFÉRENCES

¹ Statistas, Worldwide Data Created : <https://www.statista.com/statistics/871513/worldwide-data-created/>

² Rapport : 2030 Computing, Huawei

³ Veritas Technologies Projects Dark Data to Waste Up to 6.4m Tons of Carbon : <https://www.veritas.com/news-releases/2020-04-21-veritas-technologies-projects-dark-data-to-waste-up-to-6-4-m-tons-of-carbon-dioxide-this-year> *Dioxide this Year*

INFORMATIONS LÉGALES

Auteurs :

Eliott MOURIER, Partner Data Compliance & Privacy, Micropole

Pascal ANTHOINE, Data Governance Director Senior Partner, Micropole

Copyright © 2024 Micropole

Tous droits réservés. Cet ouvrage ne peut en aucune manière être reproduit en tout ou partie, sous quelque forme que ce soit, ou encore par des moyens mécaniques ou électroniques, y compris de stockage de données et leur retransmission par voie informatique, sans autorisation.

La citation des marques est faite sans aucun but publicitaire. Les erreurs ou les omissions involontaires qui auraient pu subsister dans cet ouvrage malgré les soins et les contrôles de Micropole ne sauraient engager sa responsabilité.

MICROPOL

info@micropole.com - www.micropole.com



[linkedin.com/company/micropole](https://www.linkedin.com/company/micropole)



twitter.com/groupemicropole