

**DIVISER PAR DEUX L'EMPREINTE
ENVIRONNEMENTALE DE LA DATA :**

L'ENJEU DES DARK DATA



SOMMAIRE

**Diviser par deux l'empreinte
environnementale de la data :
la chasse aux Dark Data est ouverte !** **05**

**Les données oubliées,
angle mort de la stratégie
environnementale des entreprises** **06**

**Contraintes réglementaires, absence
de compétences, manque de temps :
les origines des Dark Data** **08**

**Cyberattaques, budget stockage,
mise en conformité réglementaire :
les Dark Data coûtent cher aux entreprises** **11**

**Compta carbone des données :
quelle approche pour lutter
contre les Dark Data ? Se digitaliser,
sans oublier de se transformer** **12**



DIVISER PAR DEUX L'EMPREINTE ENVIRONNEMENTALE DE LA DATA : LA CHASSE AUX DARK DATA EST OUVERTE !

Les Dark Data, qu'est-ce que c'est ? C'est l'ensemble des données « froides » qui ne sont pas exploitées au quotidien, qui ne sont que très rarement voire jamais consultées et utilisées, qui « dorment » dans des serveurs sans que plus personne ne sache très bien pourquoi.

De quoi parle-t-on ? De documents bancaires que la régulation impose de conserver au cas où, d'archives de cessions de biens immobiliers et fonciers, de fichiers numériques médicaux, de polices d'assurance, etc.

Des données qui « dorment », donc ? Pas vraiment en fait. Stockées sur des serveurs, elles occupent de l'espace et consomment de l'énergie.

Or... 52% des données hébergées sont des Dark Data en moyenne dans le monde, 56% en France, 66% en Allemagne, 54% aux USA, etc. !

Ce ne sont pas moins de **1,3 milliard de gigaoctets**¹ de Dark Data qui sont générés par les entreprises chaque jour, soit l'équivalent d'1,3 milliard de DVD en haute définition, imaginons !

Les serveurs qui stockent la data, « chaude » ou « froide », qu'ils soient abrités dans les sous sols des entreprises ou dans des centres de données des opérateurs de Cloud, consomment de l'énergie, on le sait. Cette consommation (qui ne prend en outre pas en compte la fabrication et le cycle de vie des serveurs eux-mêmes) **compte pour jusqu'à 14% de l'empreinte carbone du numérique en France, d'après l'étude publiée par Le Green IT².**

Cela représente près de 2,3 millions de tonnes de CO₂ d'émissions, c'est énorme ! Selon cette même étude, si rien n'est fait pour réduire cet impact, celui-ci pourrait augmenter de 60% d'ici 2040 pour représenter plus de 5 millions de tonnes et autour de 0,2% du total des émissions de CO₂. Rappelons que la Stratégie Nationale Bas Carbone, menée par le Ministère de la transition écologique, prévoit une neutralité carbone à horizon 2050, avec un palier de -50% à horizon 2030.

On le comprend immédiatement : puisque les Dark Data ou les « données froides » représentent la moitié du total des données hébergées, en les stockant différemment avec des solutions « inertes », non consommatrices d'énergie, on diviserait par deux les émissions de CO₂ liées à l'hébergement de la data en France : 1,15 million de tonnes de CO₂ économisées, rien de moins.

C'est ce que ce second cahier Human Interactive vous propose d'explorer : la chasse aux données froides est ouverte !

Focus

1,3

milliard de gigaoctets
de Dark Data
sont générés
par les entreprises
chaque jour

¹ [Digital Decarbonisation](#)

² [Empreinte environnementale du numérique mondial](#)

LES DONNÉES OUBLIÉES, ANGLE MORT DE LA STRATÉGIE ENVIRONNEMENTALE DES ENTREPRISES



Focus

Chiffres clés Dark Data

175

zettaoctets de données
seront générés par l'IoT
d'ici 2025

Les chiffres liés aux Dark Data risquent d'augmenter massivement dans les années à venir et de façon inexorable, puisque 175 zettaoctets¹ de données seront générés d'ici 2025 par l'Internet des objets (IoT).

Les Dark Data, d'après différentes estimations, représentent en moyenne 52% des données stockées dans le monde et ont généré 6,4 millions de tonnes de CO₂¹ en 2020. C'est l'équivalent des émissions annuelles de dioxyde de carbone de 80 pays, ou d'une voiture qui ferait 575 000 fois le tour de la Terre. En France, elles représentent 56% de la data hébergée, soit 2,3 millions de tonnes de CO₂ par an², ou 352 876 jours de chauffage au gaz³.

L'impact environnemental des Dark Data interpelle sur la nature de ces données, et leurs processus d'accumulation

Les Dark Data désignent toutes les données inutilisées, inconnues et inexploitées d'une entreprise, générées par les interactions quotidiennes des utilisateurs – employés, clients, prestataires extérieurs – avec d'innombrables machines et systèmes. Ces fichiers et informations collectés, traités et stockés par les entreprises, mais qui ne sont plus utilisés, alourdissent l'empreinte environnementale numérique des entreprises⁴.

Elles englobent à la fois **des données techniques (fichiers de log des serveurs), mais aussi les données de géolocalisation, les emails et fichiers attachés, les données d'anciens employés, des présentations, des résultats financiers, des enregistrements d'appels clients**, etc. Ces données considérées comme trop anciennes pour avoir de la valeur, incomplètes ou redondantes, peuvent aussi se présenter dans un format de lecture incompatible avec les outils à disposition.

¹ [The benefits of reducing Dark Data](#)

² Calcul de l'empreinte carbone Dark Data :
(Taille CO₂ du numérique x % de l'empreinte carbone datacenters en phase d'utilisation) / % Dark Data

³ [Hellocarbo : tonne équivalent CO₂](#)

⁴ [Gartner – Dark Data definition](#)
[Splunk – Dark Data definition](#)

Prendre en compte l'impact environnemental des services numériques nécessite aussi de se focaliser sur la question des datas

Aujourd'hui, l'empreinte environnementale du numérique est principalement abordée via le prisme des terminaux et des équipements informatiques : allongement de leurs durées de vie, reconditionnement, utilisation d'éco-matériaux, etc.

Or, les données sont l'asset digital qui connaît la plus forte croissance. Si la crise sanitaire et la généralisation du télétravail ont décuplé le nombre de terminaux numériques

déployés (ordinateurs portables, écran, etc.), les données ont connu une augmentation non moins spectaculaire : +24% de données générées, stockées et consommées entre 2020 et 2021 et d'après l'IDC, le volume mondial de données sera multiplié par 45 d'ici 2035 pour atteindre plus de 2 100 zettaoctets.

La gouvernance des données et leur hébergement sont aujourd'hui paradoxalement occultés des pistes de réduction de l'impact environnemental. Optimiser la gestion de la consommation énergétique des centres de données et la recyclabilité des serveurs – appelés à se multiplier pour être en mesure d'héberger l'ensemble de cette data – ne saurait suffire.

Un tri des données à la Marie «Kondo» s'impose!



CONTRAINTES RÉGLEMENTAIRES, ABSENCE DE COMPÉTENCES, MANQUE DE TEMPS : LES ORIGINES DES DARK DATA

Selon une étude internationale réalisée par Splunk¹, les trois premières raisons de la prolifération des Dark Data dans les organisations sont l'accroissement de leur volume (39%), le manque de compétences permettant leur traitement efficace (34%), et le manque de ressources (32%).

Même si l'incapacité des entreprises à traiter leurs données est l'une des causes majeures avancées, les contraintes réglementaires représentent aussi un réel défi.

Ces contraintes – notamment d'archivage – génèrent également des Dark Data au sein des entreprises.

La loi française impose par exemple aux organisations d'archiver les données liées à l'exercice de leurs activités pendant une durée légale minimale. Ce délai varie selon la nature des documents et des obligations légales (2 ans à partir de la résiliation d'une police d'assurance ou 30 ans pour les contrats d'acquisition ou de cession de biens immobiliers et fonciers).

Il est intéressant de noter que les organisations dotées d'un système de gestion de contenu de type Content Services Platform (CSP)² ont une meilleure visibilité sur les données soumises à ces obligations légales, et donc un avantage potentiel quant à la gestion des Dark Data.

Faire évoluer la réglementation pourrait-il néanmoins réduire voire endiguer le flux de stockage de Dark Data ?

Non, puisque les évolutions technologiques et le réflexe humain visant à conserver, à préserver, vont générer davantage de données stockées.

Quand l'évolution technologique et le principe de précaution alimentent conjointement les Dark Data

En constante progression, les Dark Data sont directement alimentées par l'inflation des données générées chaque année. L'usage croissant des objets connectés (IOT) entraîne irrémédiablement une production massive de data. Le cabinet IOT Analytics³ estime que le nombre d'objets connectés dans le monde pourrait atteindre 21 milliards en 2025 contre 12,3 milliards en 2021.

Un autre facteur, humain cette fois-ci, la peur, alimente les Dark Data. Elle pousse à une conservation excessive des données. Pour Jérémie Cousin, Président de CIV France: *« Tout le monde veut conserver ses données depuis 2000, en cas de besoins. Les entreprises appliquent aussi le même principe en stockant des données dans nos datacenters sans les consulter pendant des années ».*

L'enjeu pour les entreprises n'est pas tant de supprimer ces données à tout prix – car certaines d'entre elles doivent être conservées pour des raisons d'archivage par exemple – mais plutôt de leur offrir un mode d'hébergement adapté.

¹ [The state of Dark data](#)

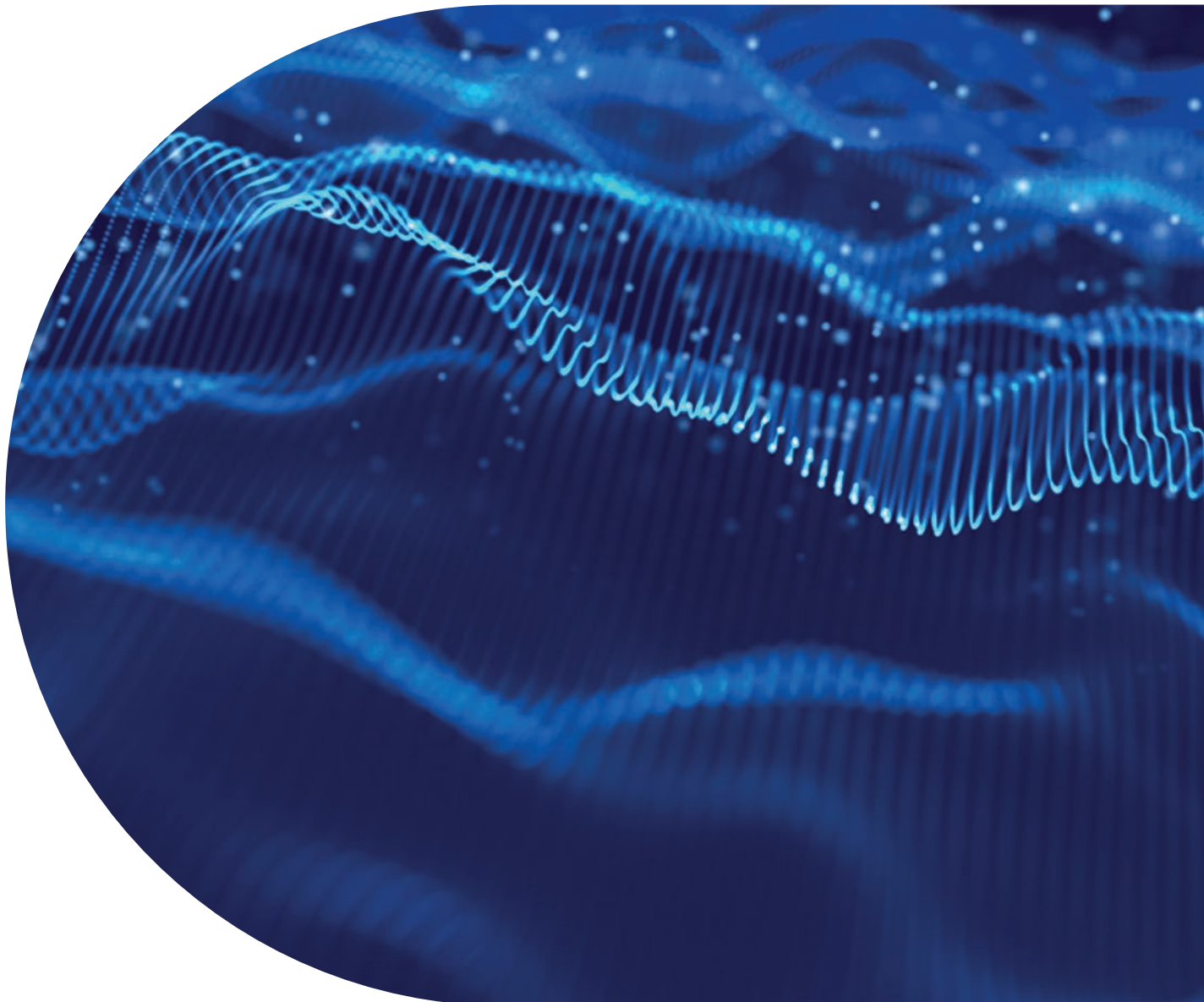
² Content Services Platform : appellation de Gartner qui remplace l'Enterprise Content Management (ECM)

³ IOT Analytics research report 2018

Comment distinguer alors les données à conserver à portée de clics, celles qui pourraient être utiles de temps en temps et celles qui seront utiles un jour, voire jamais ?

Aujourd'hui, les entreprises sont laissées seules face à cette tâche immense de tri et de classification des données. Le manque de compétences en interne et une offre limitée de ressources externes pour les guider dans ce chantier empêchent de juguler l'inflation des Dark Data.

“ Les évolutions technologiques et le réflexe humain, visant à conserver et à préserver, vont générer davantage de données stockées ”





CYBERATTAQUES, BUDGET STOCKAGE, MISE EN CONFORMITÉ RÉGLEMENTAIRE : LES DARK DATA COÛTENT CHER AUX ENTREPRISES

Les Dark Data génèrent des coûts notables pour les organisations. Elles coûteraient **2 milliards d'euros** chaque mois aux entreprises à l'échelle mondiale¹, une situation due à une augmentation significative et inutile des capacités de stockage nécessaires pour leur hébergement.

A cela s'ajoute l'impact environnemental des Dark Data qui ont généré en 2020 **6,4 millions de tonnes de CO₂** dans le monde. Une empreinte environnementale portée non seulement par les entreprises qui génèrent ces Dark Data mais aussi – et surtout – par celles qui les conservent et qui les traitent.

En effet, que ce soient les Cloud providers, les entreprises d'analyse et de gestion de données ou encore les éditeurs de logiciels, tous subissent les conséquences des Dark Data.

Le groupe Tessi, par exemple, en offrant des solutions de stockage à ses clients, se retrouve à devoir assumer l'hébergement de ces données. Quelle proportion des 2,3 pétaoctets de data clients conservés par Tessi doit-on d'ailleurs classer en tant que Dark Data ?

Au-delà de ces impacts financiers et environnementaux pour les entreprises, d'autres facteurs peuvent entraîner des conséquences non-négligeables et sensibles pour les entreprises.

Les Dark Data, facteurs de risques pour les entreprises

Selon une étude réalisée par HelpNetSecurity², on retrouve **42% d'informations confidentielles, 1% d'informations personnelles sensibles, et 9% d'informations personnellement identifiables au sein des Dark Data**. Un risque réel de **cybersécurité** qui pèse donc sur les entreprises en ne traitant pas ces données.

En effet, celles-ci ne sont pas uniquement composées de data inutile. Dans la masse de données hébergées par les entreprises se trouvent certaines informations d'importance stratégique. Un piratage de ces fichiers non classifiés et stockés parmi tant d'autres pourrait entraîner des conséquences majeures pour les entreprises et leurs employés (ransomware, vente de données sur le dark web, etc.).

Un risque qui appelle à la mise en place d'une stratégie de gouvernance des données. Une approche validée par Jean-Philippe Balivet, Directeur des activités CSP de Tessi France : *« Il faut sensibiliser les entreprises à la gouvernance de la donnée et de la vie du document. Pour réduire le risque de cybersécurité, il faut déjà analyser et évaluer l'importance de la donnée en amont de son stockage »*.

Cette nécessité d'encadrer la data et son utilisation met en lumière un autre risque auquel font face les entreprises : **la conformité en matière de data personnelle**. Réglementée par le RGPD, certaines données personnelles doivent être conservées par les entreprises pour une durée limitée (3 ans en général). Au-delà de cette limite, les données doivent être supprimées par les entreprises. Or, ces données sont rarement classifiées par les entreprises, se perdant au milieu des Dark Data, et ne sont généralement pas supprimées au bout des 3 ans. Une négligence qui peut faire tomber les entreprises sous le coup de la loi et remet en perspective l'impératif de classification des données, géré par les plateformes de services de contenu (CSP).

Focus

Les Dark Data coûteraient

2 M^{rds}€

chaque mois
aux entreprises

¹ IDC Market Analysis Perspective: Worldwide Managed Support Services 2020

² [Etude HelpNetSecurity](#)

COMPTA CARBONE DES DONNÉES : QUELLE APPROCHE POUR LUTTER CONTRE LES DARK DATA ? SE DIGITALISER, SANS OUBLIER DE SE TRANSFORMER

Répondre au défi de la transition écologique posé par les Dark Data impose de trouver des solutions qui permettent à la fois de les diagnostiquer et de les trier, mais surtout d'accompagner les entreprises dans ce tri.

L'objectif d'Innovation & Trust visant à réduire de 15% à 20% le volume des Dark Data au niveau des entreprises d'ici cinq ans passe par la mise au point d'un dispositif dédié à la mesure et à la gestion de ces « données froides » : « **EKODATA** » by Innovation & Trust.

Plateforme de diagnostic, de tri et de classification des Dark Data, « **EKODATA** » by Innovation & Trust vise à maintenir les activités de stockage dans un cadre écoresponsable et à apporter une aide aux entreprises manipulant des volumes importants de données.

La méthode Innovation & Trust : une compta carbone pour chasser les Dark Data des serveurs

« **EKODATA** » by Innovation & Trust s'appuie sur plusieurs fonctionnalités innovantes.

Un calculateur de Dark Data – Track ma Data – permettra à une entreprise de mesurer et d'assurer un suivi de l'estimation de son impact carbone, énergétique et d'éventuelles économies en matière de coût d'hébergement, sur la base du volume de données stocké par l'organisation.

Le calculateur pourrait ressembler à celui proposé par le [Digital Decarb](#) :

Calculateur du coût en carbone des données

Veuillez indiquer le nombre d'employés :

50

Calculer les coûts

Coût du carbone des données pour tous vos employés :

2 295 gb

Données générées par jour pour tous les employés

5 tonnes de CO₂

La quantité de CO₂ générée quotidiennement par vos nouvelles données

L'empreinte carbone est similaire à celle d'un vol de Londres Heathrow à New York

5/jour !

550 800 gb

Données générées par an pour tous les employés

1 102 tonnes de CO₂

La quantité de CO₂ générée annuellement par vos nouvelles données

L'empreinte carbone est similaire à celle d'un vol de Londres Heathrow à New York

1 281/an !

Dark Data Cost

En utilisant l'Intelligence Artificielle (IA) et le Machine Learning, cette plateforme de gestion des Dark Data scanne, identifie, étiquette et classe les informations en s'assurant que les données sensibles ou à risque sont protégées. Elle peut également proposer un archivage des données jugées importantes mais moins utiles dans un système de stockage « glacier », et recense les fichiers liés à le RGPD avec la mise en place d'alertes ou notifications déclenchées à l'approche de l'expiration des autorisations reçues.

Cette plateforme est destinée aux entreprises traitant de gros volumes de données, quelle que soit leur taille ou leur secteur d'activité (retail, santé, finances, assurance, etc.).

Il s'agit de la première solution mise sur le marché qui adresse spécifiquement les Dark Data et qui offre une ressource aux entreprises pour les accompagner dans la gestion et le tri de leurs données.

Innovation & Trust prévoit sa mise en service pour le premier trimestre 2023.

Une approche responsable et engagée au service des clients

Chez Innovation & Trust, nous avons à cœur de conjuguer développement technologique et croissance économique en tenant compte de l'impact environnemental de nos activités. Engagés dans un programme numérique responsable, nous développons des solutions pour faire converger transition digitale et transition environnementale et souhaitons accompagner nos clients dans cette démarche.

Toutes nos solutions sont d'ores et déjà développées dans une approche responsable: une solution bien codée dès la première ligne de code est une solution éco-conçue !

Innovation & Trust est une filiale du groupe Tessi :

Les engagements du groupe Tessi pour un Numérique Responsable

- Tessi a rejoint, depuis 2020, l'initiative Planet Tech'Care auprès de NUMEUM (ex Syntec Numérique) <https://planet-techcare.green>
- Tessi signataire du manifeste NUMEUM pour l'inclusion et la reconversion des femmes dans les métiers du numérique aux côtés de plus de 150 autres entreprises engagées
- Tessi membre de l'Institut du Numérique Responsable et signataire de la charte Numérique Responsable depuis Février 2022



Innovation&trust



Innovation&trust

Innovation&trust

Immeuble Ileo
27 - 33 Quai Alphonse le Gallo
92100 Boulogne-Billancourt
www.tessi.eu