

Tasm^{ane}

Green IT et sobriété

numérique :

et si on commençait par connaître et mieux
gérer ses données ?

LA PAROLE

Valérie VIEILLEFONT

Manager

Soha BENCHEKROUN

Consultante



FEVRIER 2022



LA PAROLE

Valérie VIEILLEFONT & Soha BENCHEKROUN

‘ La data, une dimension peu évoquée de la sobriété numérique ’

Terme en vogue pour des raisons environnementales évidentes, la « sobriété numérique¹ » vise à réduire l’empreinte écologique du secteur, qui représente aujourd’hui **3 à 4 % des émissions de gaz à effet de serre (GES)** dans le monde² et est un très grand consommateur de ressources rares, non renouvelables. Plusieurs leviers sont possibles, à commencer tout simplement par la réduction de l’usage des objets numériques et l’allongement de leur durée de vie des data centers.

Moins souvent évoquée, la gestion des données pourrait aussi être une piste pour proposer des solutions IT plus sobres d’un point de vue environnemental. Une approche responsable pourrait-elle être compatible avec une meilleure façon d’exploiter tout le potentiel des données d’une entreprise ?

Du stockage à l’analyse prédictive, les promesses du big data et de l’intelligence artificielle séduisent tous les secteurs économiques. Le Medef lui-même vient de publier en septembre 2021 un plaidoyer pour le traitement par les entreprises françaises de la data comme un véritable atout³.

Pourtant, de plus en plus de voix s’élèvent pour pointer l’impact écologique croissant du « big data » en tant que contributeur aux émissions de gaz à effet de serre. Les data centers sont énergivores à plusieurs titres : énergie nécessaire pour le fonctionnement, pour les systèmes de refroidissement des infrastructures et pour le traitement des données. Certes, les économies d’échelle et l’amélioration de l’efficacité énergétique ont permis, dans une certaine mesure, de limiter l’augmentation des émissions liées à l’engouement pour ce nouvel or noir. Mais pour l’avenir, les avis des experts divergent. Des projets de réutilisation de la chaleur produite par les data centers existent : chauffage de piscine publique ou de bâtiments à proximité, nouvelles solutions de refroidissement, etc. Mais ces pistes ne suffiront vraisemblablement pas à maintenir le niveau d’émission actuel et encore moins à le faire diminuer, d’autant qu’un effet rebond⁴ est constaté comme pour beaucoup d’innovations technologiques.

¹ Concept apparu il y a une dizaine d’années en France sous l’impulsion de Frédéric Bordage, ingénieur et fondateur de GreenIT.

² Mission d’information sur l’empreinte environnementale du numérique - SENAT - juin 2020

³ La stratégie data comme source de valeur pour l’entreprise, Septembre 2021.

⁴ L’effet rebond désigne le phénomène observé lorsqu’une réduction des facteurs limitant l’usage d’une ressource entraîne, de façon paradoxale, une augmentation de sa consommation. Par exemple, les progrès énergétiques réalisés dans certains domaines ne s’accompagnent pas d’une baisse globale de la consommation mais au contraire d’une augmentation, la ressource libérée étant orientée vers d’autres usages.



LA PAROLE

Valérie VIEILLEFONT & Soha BENCHEKROUN

Après avoir rempli le « data Lake », peut-être faut-il le vider (au moins quelque peu) ?

Il ne s'agit pas de « jeter le bébé avec l'eau du bain ». Le big data peut, dans une certaine mesure, servir les enjeux environnementaux : les exemples de recours à ce type de technologies pour construire des systèmes résilients (villes, bâtiments, ...) ou adapter les réactions face à certains phénomènes naturels tels que les pics de pollution ou les crues se multiplient.

En réalité, **la problématique est moins le phénomène du « Big data » et de la data science que deux assertions fondamentales qui les sous-tendent :**

- Plus de données = plus de valeur,
- La systématisation du principe consistant à "conserver" les données malgré l'absence de use case à date (car on ne sait jamais, « ça peut toujours servir un jour »). Une des caractéristiques de la data science étant d'être exploratoire et de ne pas systématiquement savoir à l'avance l'emploi auquel on destine les volumes énormes de données stockées.

Ces deux principes doivent être battus en brèche et mieux analysés pour engager pleinement la filière data vers plus de sobriété. Dans la droite ligne du mouvement "slow food" né au début des années 1980, qui a été décliné pour dans d'autres secteurs d'activité avec le "slow travel" et le "low tech" (...) **pourquoi ne pas développer le "slow data" en intégrant les principes de sobriété et d'hygiène dans le management des données** et en partant du principe que **volume et conservation systématique se sont pas une fin en soi ?**

Des études montrent régulièrement qu'une part importante des données stockées par les entreprises ne sont pas exploitées, appelée « **Dark data**⁵ » - cette fraction peut représenter jusqu'à 80% des informations stockées. Un discours habile consiste à souligner l'importance de les exploiter plus encore. Pourtant, nous pourrions aussi poser la question inverse plus abruptement : l'organisation en a-t-elle vraiment besoin ? Toutes ces données sont-elles porteuses de valeur ? Bien que tout le monde ait en tête des exemples de jeux de données pas/mal exploités, nous savons aussi que des pans substantiels de données n'apportent que peu de valeur une fois le processus métier (ou technique) terminé.

⁵ Les Dark Data représentent toutes les données que les organisations possèdent mais qu'elles n'exploitent pas.



LA PAROLE

Valérie VIEILLEFONT & Soha BENCHEKROUN

Quelques pistes simples issues de principes basiques du data management pour y parvenir

En regardant de plus près les principes de gouvernance des données⁶, nombre d'entre eux servent à la fois les enjeux de maximisation d'apport de valeur mais également les enjeux environnementaux :

- Un des points fondamentaux concerne **la gestion du cycle de vie des données**. Appliquée à une large échelle, cette pratique permet de préciser et mettre en œuvre des actions essentielles pour limiter la volumétrie : définir une durée maximum de stockage, appliquer des mécanismes de purge et d'archivage, éviter les copies de jeux complets de données dans leur format d'origine (souvent déversés dans un datalake ... ayant tendance à devenir un Data Swamp⁷).
- **La connaissance des données** est une discipline en plein essor, dont le catalogage est l'illustration la plus fréquente. Elle a pour effet principal de réduire le time-to-market des projets et le temps de recherche des informations. Il permet également de renforcer leur appropriation par les équipes métiers et IT. Cette connaissance du patrimoine data, essentielle pour participer à la création de valeur de l'entreprise, peut aussi être utilisée pour débusquer les données qui en ont moins (voire pas) et qui ne méritent pas une persistance outre mesure dans le SI.
- **Le développement des approches pilotées par les usages** participe à la conservation et la mise à disposition des seules données d'intérêt. **Le smart data**, utilisé dans le monde du marketing, prône la pertinence et la qualité plutôt que la quantité. Dans une de ses études, Gartner estime que 70% des organisations ont la volonté de se tourner du Big vers le small data⁸.
- Enfin, soulignons que l'un des principes centraux du RGPD est celui de **la minimisation des données**, qui prévoit de limiter les données à ce qui est nécessaire et pertinent compte tenu des finalités et de la durée du traitement.

⁶ Par exemple ceux du référentiel DAMA (DAta Management Association)

⁷ « Marécage de données » : lac de données inaccessibles aux utilisateurs concernés, ou de peu de valeur.

⁸ Gartner Says 70% of Organizations Will Shift Their Focus From Big to Small and Wide Data By 2025. Mai 2021.



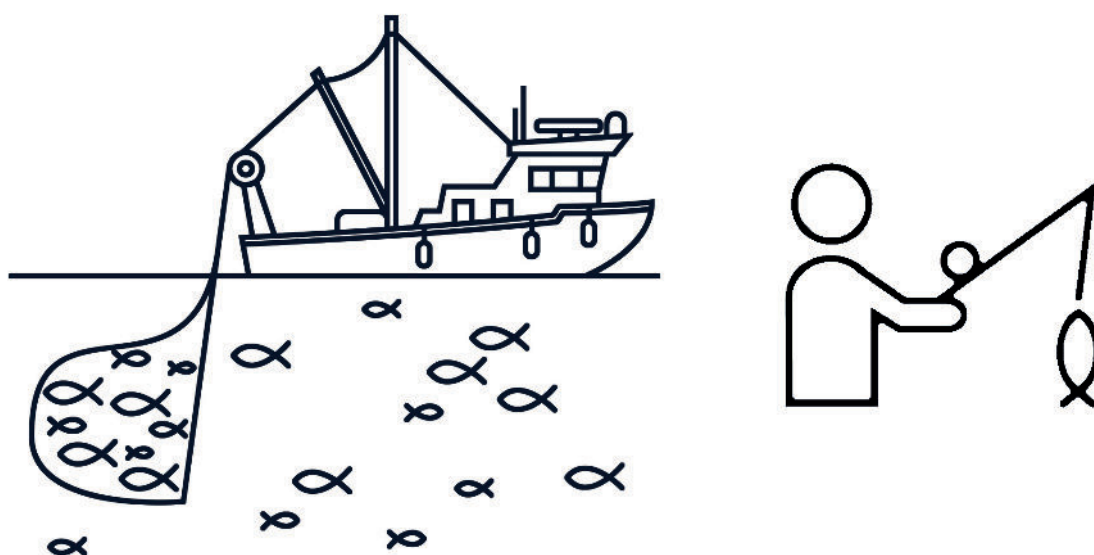
LA PAROLE

Valérie VIEILLEFONT & Soha BENCHEKROUN

Outre la gouvernance, certaines techniques concourent également à la limitation du volume de données gérées, et donc limitent les impacts environnementaux, dont voici deux exemples :

- **La compression de données**, souvent utilisée dans le cadre de collecte à partir de capteurs. Il s'agit, par exemple, au niveau d'un capteur, de calculer et stocker une valeur moyenne plutôt que les valeurs unitaires de N relevés consécutifs.
- **La virtualisation de données**, qui permet d'accéder et de visualiser des données distribuées dans de multiples data stores hétérogènes. Plutôt que d'effectuer des opérations d'extraction et de transformation sur les données de façon physique à l'aide de briques classiques (ETL), les serveurs de Data Virtualization effectuent l'extraction, la transformation et l'intégration virtuellement. De fait, cette technique réduirait à son tour les redondances de stockage et contribuerait à une rationalisation des volumes de données dans la mesure où seules les métadonnées sont stockées.

Ces exemples ne sont pas limitatifs et montrent que les entreprises et leur DSI ont intérêt à intégrer la gestion des données dans leurs initiatives de sobriété numérique, voire pour certaines à en faire un véritable marqueur de culture d'entreprise.



Plus de données = plus de valeur ?



En savoir plus

Vous souhaitez poursuivre les échanges ?



Contactez Valérie VIEILLEFONT :

valerie.vieillefont@tasmane.com