

Les impacts environnementaux et sociétaux des données



Un défi pour l'avenir



JRES - Mai 2022



mis à disposition selon les termes de la [licence Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).



Didier Mallarino, Sylvie Le
Bras & Cyrille Bonamy



- Création **EcoInfo** en 2006 ; Groupement De Service (GDS) en 2012.
- Soutenu par le **CNRS** ; 2 instituts : **INS2I & INEE**
- Environ **60** membres de l'ESR en France qui travaillent autour d'un objectif commun :



Agir pour réduire les impacts (négatifs)
environnementaux et sociétaux des TIC

Agir

A ACTION
C CHANGES
T THINGS

Contexte

CIJC 2025

ANTHROPOCENE

Les Indicateurs sont largement dans le rouge
une situation potentiellement délétère pour notre espèce

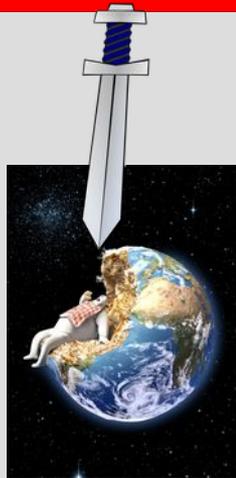
Jour du dépassement

Global = 1,7 planète

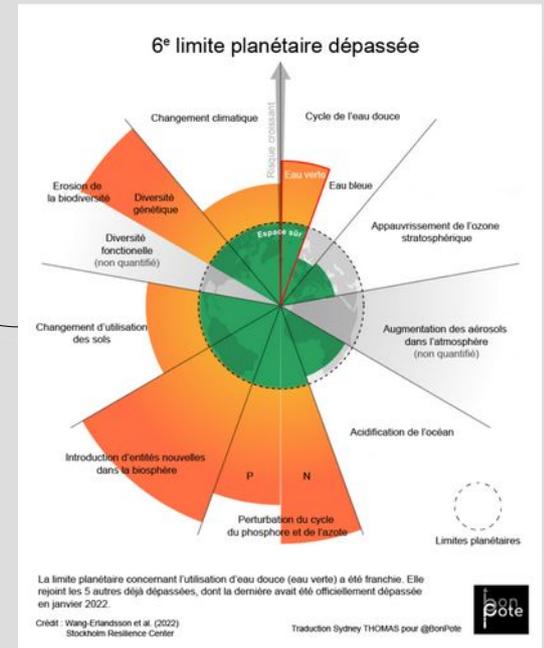
Qatar, 9 février

Vietnam, 21 décembre

France 5 Mai 2022



- Changement climatique
- Biodiversité une espèce sur huit risque de disparaître à brève échéance & sixième extinction de masse
- Destruction des sols et des habitats
- Pollutions (eaux, sols, atmosphère)
- Modifications des cycles biogéochimiques & du cycle de l'eau douce



JRES 2022 - Marseille

Limite Planétaires (Planetary Boundaries) - Rockström & al (2009) et Steffen & al (2015) & Bon Pote ;
<https://bonpote.com/la-6e-limite-planetaire-est-franchie-le-cycle-de-leau-douce/>

Le numérique

Focus données

une approche **inhabituelle** de l'aspect le plus **immatériel** qui est pourtant **la raison d'être du numérique et de ses impacts.**

TIC : ~ 4 % des GES
et une croissance forte des usages
(+6% an)



« Le numérique » s'inquiète enfin de ses **impacts sur l'environnement** et ne se considère plus uniquement comme faisant partie de la **solution** mais comme également partie du **problème posé à l'espèce humaine.**
Comme les autres secteurs, il doit « **faire sa part** » de réduction

Transports ~ 35% GES



Aviation ~ 4% GES

La donnée : définition

La donnée en informatique est la représentation numérisée d'une information physique ou mentale

La donnée : utilisation

- Construction du **savoir**
- & des **connaissances**

La donnée en informatique est la représentation numérisée d'une information physique ou mentale

La donnée : éthique

Un pouvoir décuplé par les outils informatiques et notamment l'IA

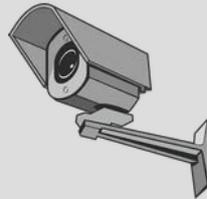
- Construction du **savoir**
- & des **connaissances**
- **Outil de pouvoir**

dataveillance & datapulation

La donnée en **informatique** est la représentation numérisée d'une information physique ou mentale



https://fr.wikipedia.org/wiki/La_Linea



La donnée : utilisation et éthique

Un pouvoir décuplé par les outils informatiques et notamment l'IA

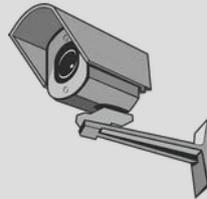
- Construction du **savoir**
- & des **connaissances**
- **Outil de pouvoir**

dataveillance & datapulation

La donnée en **informatique** est la représentation numérisée d'une information physique ou mentale



La Linéa d'Oswaldo Cavandoli



- **Pouvoir politique, financier et commercial**
 - Suggestions & influence
 - **manipulations** psychologiques (tirant parti de nos biais cognitifs: e.g : achat impulsifs vs agir rapidement)
 - Fake news (guerre informationnelle)
 - Société de **surveillance** (Chine par exemple)
- **Utilisation du temps de cerveau disponible → Addictions**
- « Bulle de filtres » (isolement) & **désinformation**
- **Analyse des états émotionnels & profils psychologiques**
 - **250 likes** suffisent à établir un **profil** de **personnalité** précis (type « OCEAN »)

La donnée : quelques clés

- **L'acquisition** est **coûteuse** écologiquement (& financièrement)
- La **donnée** est **précieuse et/car unique** (trace d'un instant révolu)
- La donnée « brute » est **inutilisable** et doit être associée à des **méta données** pour la rendre **utile, utilisable, pérenne, échangeable**
- La « valeur » de la donnée reste **subjective**, **dépendant de l'usage et de l'utilisateur, mais**
 - **L'histoire prouve que la connaissance issue de ces données apporte de nombreux avantages en termes de pouvoir et de richesses**
 - **L'appétence pour les données** est immense et la plupart des entreprises et des états se **battent** pour récolter de la donnée

La donnée : leur forte croissance

Usages



IOT (30 milliards en 2023), images sons & vidéo, santé, IA

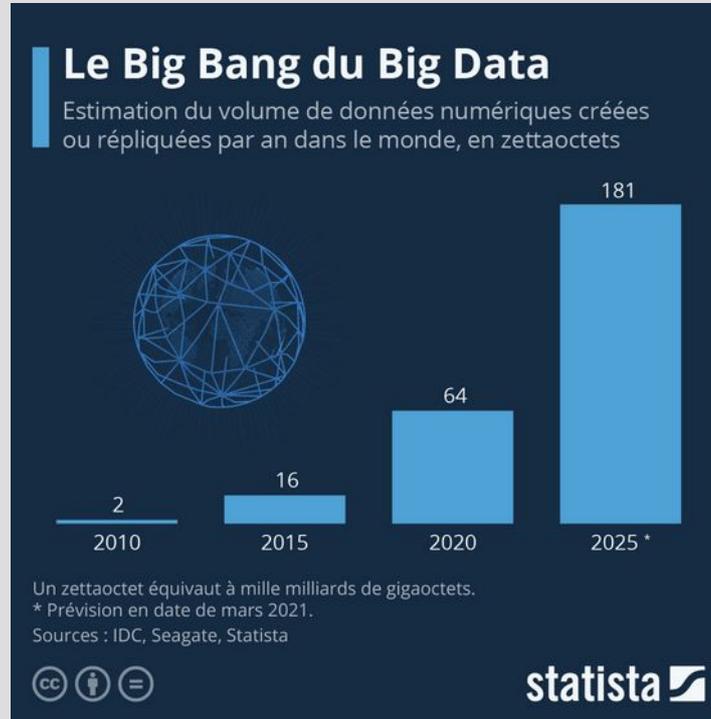
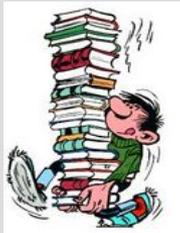


En 2020, 60 Zo
« créés » &
~ 1 % stockés

1 Zo = 1000 Eo = 1500 milliards de CDROM (soit une pile de 1,5 milliards de km)

Transport

5G, fibre optique



Stockage

- Capacité de **stockage** mondiale ~ **7 Zo**
- Prévission de croissance :
 - ✓ **Données** générées :
+**40 %/an** à 5 ans
 - ✓ **du stockage** à 5 ans :
+**20 %/an**



Data Cloud & **Stockage**

Ce qui pose quelques problèmes pratiques

- **Trouver ou retrouver des données**
 - ✓ catalogage, méta données & thésaurus thématiques, identification unique et pérenne
- **Gérer la variété des formats et des contenus**
 - ✓ standards interopérables, open data
- **S'assurer de la qualité et des accès aux jeux de données (licences)**
 - ✓ Métadonnées, qualification de ses données en suivant les standards du domaine
- **Taille & nombre de fichiers**
 - ✓ normalisation, standards interopérables, open data, outils standards
- Mais aussi, en termes de **transport** (écoutez la présentation de Marion), de **stockage** et de **traitement de ces volumes croissants**

Coût du stockage

- Ordre de grandeur du **coût carbone du stockage** : l'étude menée au sein de GRICAD (Infrastructure de calcul intensif et de données de Grenoble) estime à l'année un coût de **12 à 37g CO₂ e par Go stocké**
- Cette étude pointe **3 axes forts pour minimiser l'impact du stockage** (facteur 6 possible d'amélioration sur ces axes):
 - Le bon dimensionnement pour s'appuyer sur les besoins réels
 - La durée de vie du matériel
 - Le mix énergétique utilisé pour produire l'électricité
- D'autres axes (facteur 2 environ) :
 - Rendement du datacentre (PUE)
 - Télétravail

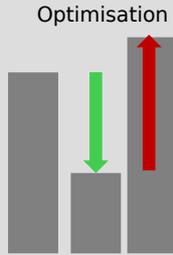


Estimation de l'empreinte carbone du stockage de données
Guillaume Charret, Alexis Arnaud, Françoise Berthoud, Bruno Bzezniak,
Anthony Defize, Yves Delay, Fabien Drago, Gabrielle Feltin, Nicolas Gibelin,
Gaël Guennebaud, et al.

► To cite this version:

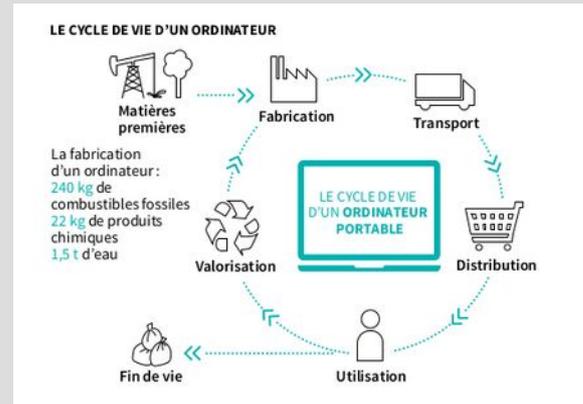
Guillaume Charret, Alexis Arnaud, Françoise Berthoud, Bruno Bzezniak, Anthony Defize, et al. Estimation de l'empreinte carbone du stockage de données. [Rapport de recherche] CNRS - GRICAD. 2020. hal-03573790

Et un cortège d'effets très concrets

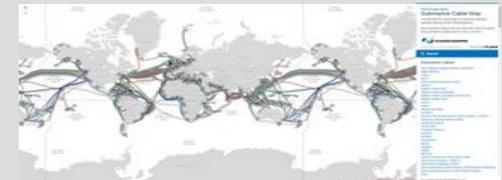
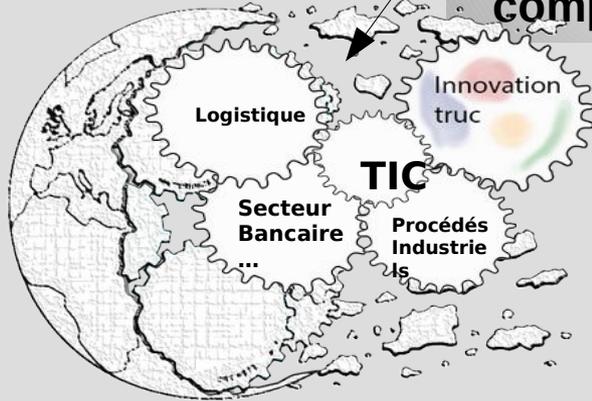


- Effets Directs
- Effets Indirects
- Effet Rebond
- Accélération, dépendance & complexité

Il n'y a pas que les GES qui sont une menace



Chiffres (2003) du cycle de vie de l'étude de « Kuehr et Willians »



Agir



Mais où, et comment agir quand on prend les choses sous l'angle de la donnée ?

Pour éviter ça



Source : Album Idées noires ;

Dessin de Franquin :

[https://fr.wikipedia.org/wiki/Id%C3%A9es_noires_\(bande_dessin%C3%A9e\)](https://fr.wikipedia.org/wiki/Id%C3%A9es_noires_(bande_dessin%C3%A9e))

Agir



Mais où, et comment agir ?

Pas de solutions magiques

Où ? cycle de vie de la donnée

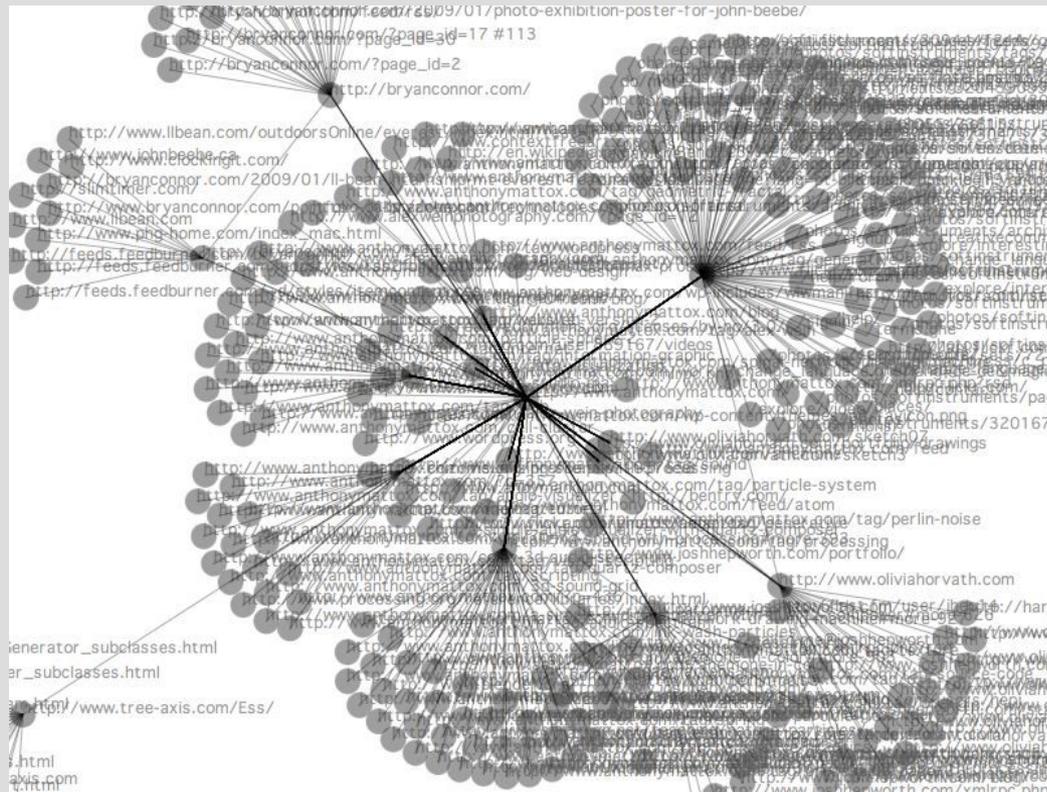
La donnée est la raison même de l'existence du matériel et du logiciel qui permettent l'acquisition, le traitement, le stockage et l'archivage ainsi que la réutilisation ou la destruction de ces données, en clair, son cycle de vie. On peut donc agir à pleins d'endroits :

On doit agir à chaque étape du cycle de vie de la donnée

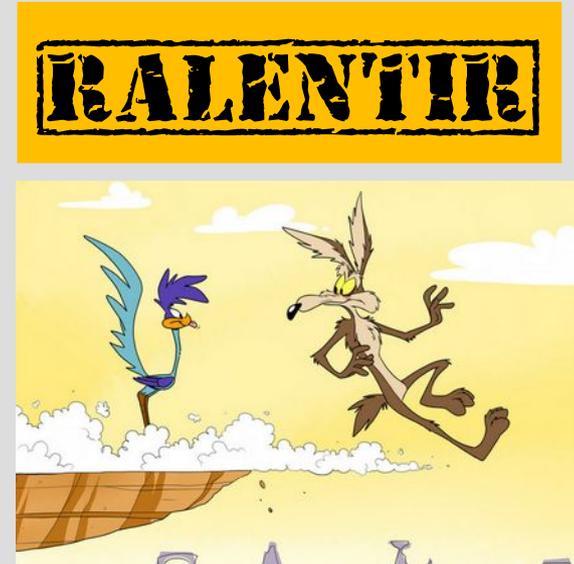
Se fixer quelques objectifs

- **Éviter** de participer à l'inflation permanente des données
- **Ne conserver que** des données **utiles** et **réutilisables** (donc, des données qualifiées, identifiées, accessibles, interopérables). Penser « **durée de vie** » des données.
- **Minimiser** le transport lors de l'usage
- Choisir les technologies de **stockage en fonction de l'usage**

Et face à la complexité



Agir

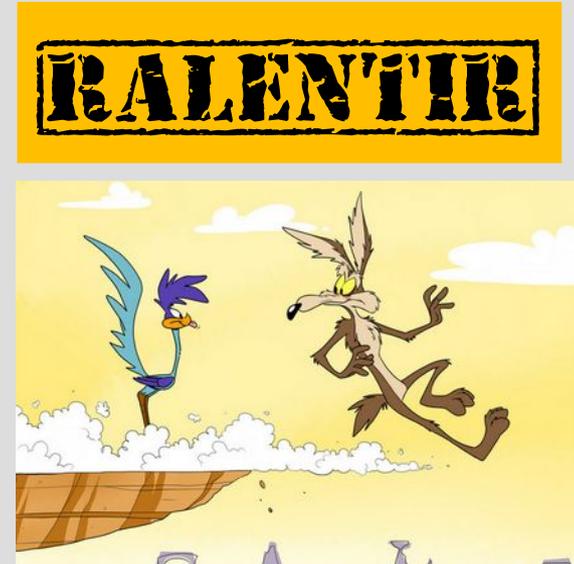


Source : [Bip Bip](#) et le coyote, Studio Warner, [Chuck Jones](#)

Agir



REFLECTIR



Source : [Bip Bip](#) et le coyote, Studio Warner, [Chuck Jones](#)

Agir

LES 7 R



REFLECHIR



RALENTIR



<https://profsentransition.com/sobriete-numerique/>

Pistes pratiques

Réfléchir

- **Organiser son architecture** pour tenir compte des flux de manipulation des données
 - Construire une architecture en fonction des volumes et des coûts énergétiques
 - Décentraliser ? Centraliser ?
- Penser **simplicité** et **résilience : comment faire sans numérique ou un numérique dégradé ?**
- **Normaliser & standardiser pour pérenniser** : utiliser des outils et des formats **ouverts** (fichiers, traitement)
- Mieux **stocker** et **archiver** ⇒ **technologies sobres en fonction de l'usage**
 - Déduplication ? SSD vs HDD ? (stockage, backups, données chaudes),
 - Bande ? Stockage hors ligne ? (données froides)

Pistes pratiques

Ralentir

- le **gain temporel** potentiel du traitement de mes données et l'importance de mon projet rendent-ils acceptable le **coût environnemental** de fabrication de ce nouveau matériel ?
- Prendre le temps de **sécuriser** et **qualifier** ses données (leur coût d'acquisition justifie à lui seul cet investissement humain pour permettre une **réutilisation** et la **reproductibilité scientifique**)

Pistes pratiques

Réduire

- **Fermer** le robinet des données inutiles
- **Ne pas acquérir (ni conserver)** de données inutilisables et s'interroger sur la durée de vie des données
- **Relocaliser** ses données au plus proche de l'usage pour **limiter le transport**

Pistes pratiques



Réutiliser : Ouvrir les données

« Les données de la recherche sont la matière première de la connaissance. Les partager, c'est ouvrir de nouvelles perspectives scientifiques » - Frédérique Vidal, ministre de l'Enseignement supérieur, de la Recherche et de l'Innovation (Juillet 2018) - Site Ouvrir la science : <https://www.ouvrirelascience.fr>

Approche FAIR : Findable, Accessible, Interoperable, Reusable



Pistes pratiques

Refuser

- **La course en avant** à un nouveau hardware, **l'obsolescence** « commerciale » voire « écologique » : **Faire durer**
- S'interroger sur **le sens et les valeurs des projets auxquels on participe** eu égard aux enjeux ?
- Protéger sa vie privé : **Data poisoning** ?



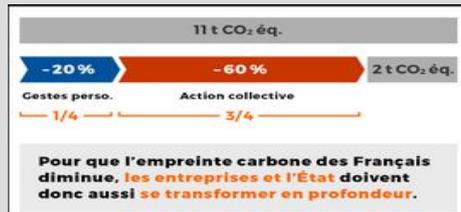
- TU VOIS, PETIT, SUR CETTE CLÉ, J'AI UNE PETITE FORTUNE EN CRYPTOMONNAIES ET EN NFT, HÉHÉHÉ!

Conclusion : élargir sa pensée

Refuser, Réduire, Réutiliser, Recycler, Rendre à la terre



FAIR(E) Sa Part



Ralentir pour prendre le temps de réfléchir ...



Et réfléchir pour arrêter de raisonner en « silos » : pensée complexe



Sobriété, low tech et décroissance



Science ouverte



JRES 2022 - Marseille

Merci :-)



« Le peu qu'on peut faire, le très peu qu'on peut faire, il faut le FAIR(E) »

Théodore Monod



Réseaux et guides

- **EcoInfo** : AGIR sur les données de la Recherche :

→ <https://ecoinfo.cnrs.fr/2021/03/01/agir-sur-les-donnees/>



- Réseau **SIST** de l'INSU :

→ <http://sist.cnrs.fr>



- « *Guide de bonnes pratiques sur la gestion des données de la recherche* »

→ <https://mi-gt-donnees.pages.math.unistra.fr/guide/00-introduction.html>

