



BIG DATA :

UNE CRITIQUE SOCIO-ÉCOLOGIQUE RADICALE

Défi Big Data (2A) Développement Durable &
Responsabilité Sociale et Environnementale

Jérémy DÉsir

Mars 2021

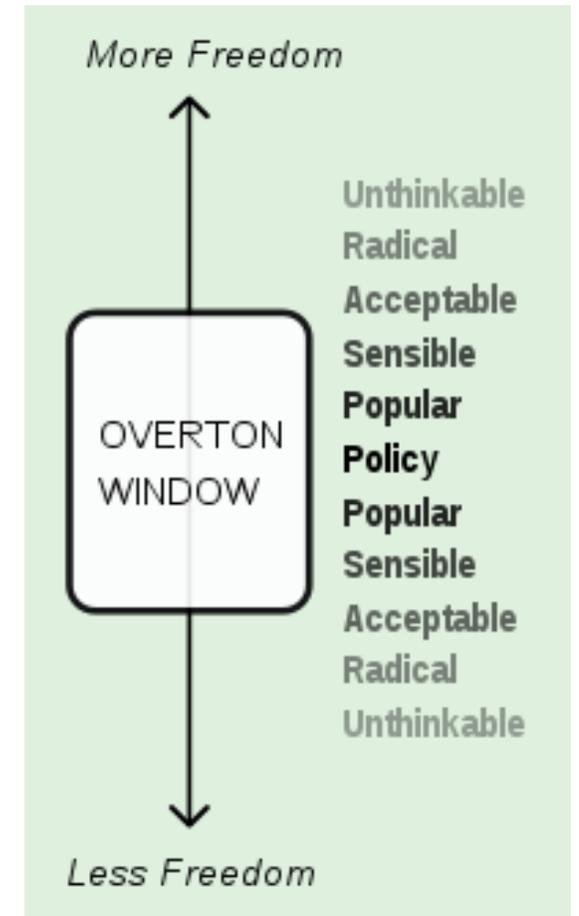
Quel est l'objectif ?



Choquer et impressionner.

Quel est l'objectif ?

- Overton décrit une carte des idées du « plus libre » au « moins libre » concernant l'action du gouvernement (ou toute action publique) sur un axe vertical.
- Comme la fenêtre change de taille ou se déplace, une idée à un endroit donné peut devenir plus ou moins politiquement acceptable.
- La fenêtre d'Overton permet d'identifier les idées définissant le domaine d'acceptabilité de certains objectifs politiques :
 - Les partisans de politiques **en dehors de la fenêtre** cherchent à persuader ou éduquer l'opinion publique afin de déplacer et/ou d'élargir la fenêtre.
 - Les partisans **dans la fenêtre** soutenant les politiques actuelles, ou similaires cherchent à convaincre l'opinion que les politiques situées en dehors de la fenêtre doivent être considérées comme inacceptables.
- Essayons, ensemble, d'élargir cette fenêtre... vers le haut.



Plan

- **INTRODUCTION**
- **ÉTAT DES LIEUX**
 - Qualité de vie
 - Environnement
 - Démocraties et libertés
- **SUGGESTIONS**
 - Déconstruire
 - Quelques notions d'OSINT
- **CONCLUSION**

INTRODUCTION



Data Scientist: *The Sexiest Job of the 21st Century*

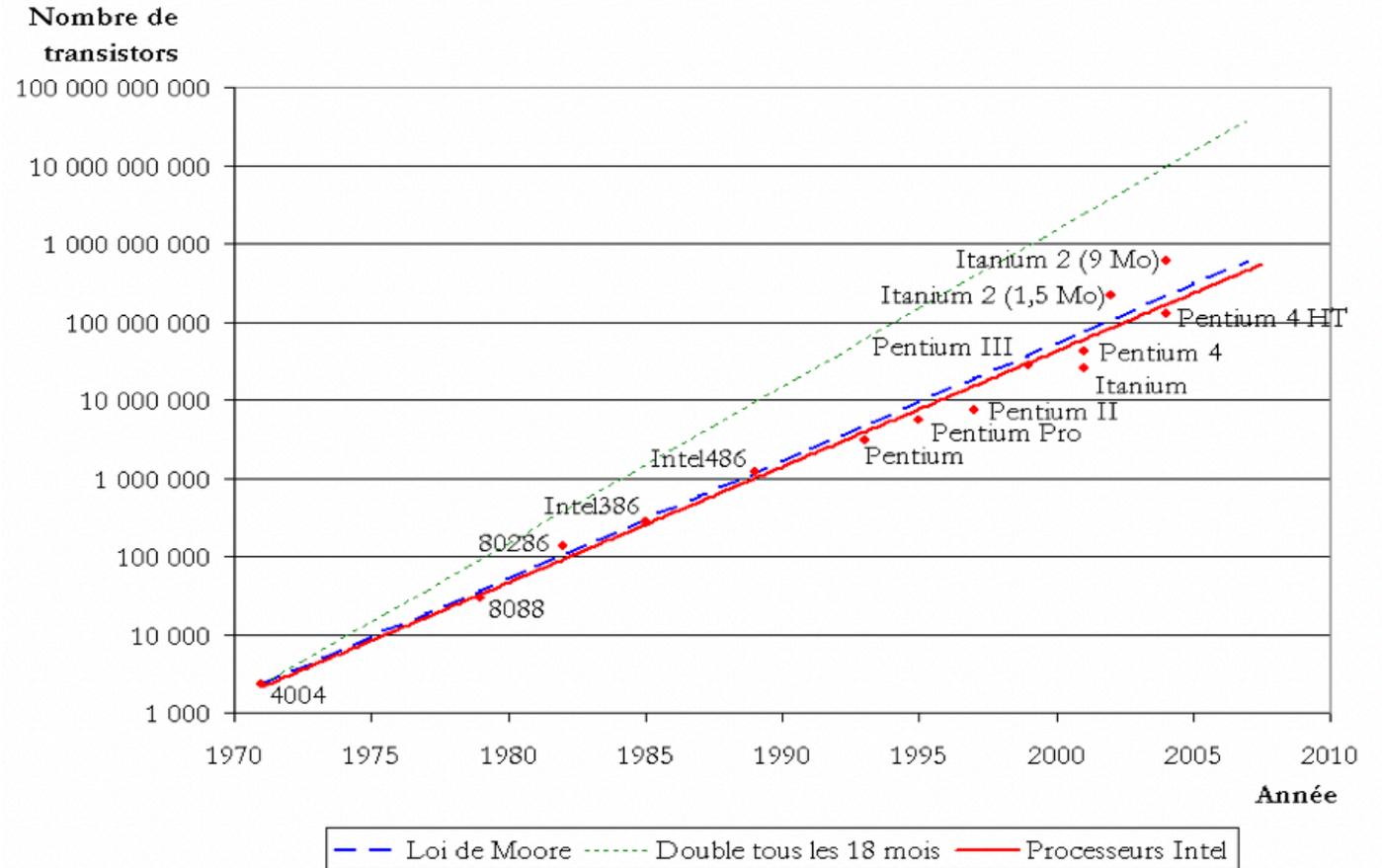
Meet the people who can coax treasure out of messy, unstructured data.

by Thomas H. Davenport and D.J. Patil

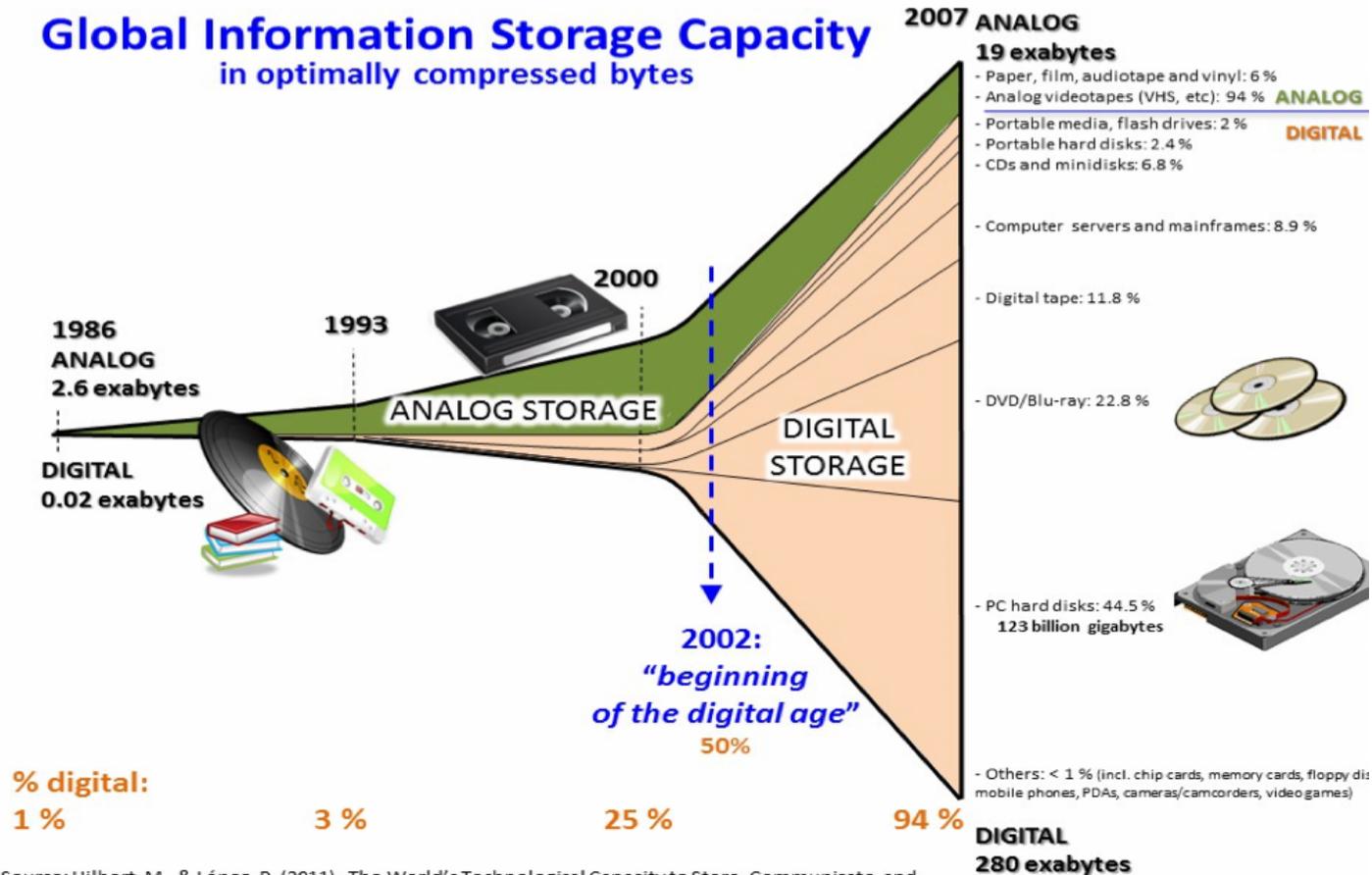
When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

Une puissance de calcul exponentielle

- **Loi de Moore** : le nombre de transistors des microprocesseurs, c'est-à-dire la puissance de calcul des machines, double tous les deux ans
- **Contraintes physiques** : dissipation thermique complexe avec des transistors de moins de 10 nanomètres)
- **Malgré ces contraintes** : Calcul parallèle, processeurs graphiques, cloud computing... et bientôt informatique quantique; fournissent un sursis à l'exponentiation de la puissance de calcul



L'essor du stockage numérique et des flux d'informations



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. <http://www.martinhilbert.net/WorldInfoCapacity.html>

- Développement tout aussi vélocité des mémoires informatiques
- Et donc du stockage et des flux d'informations numériques.
- Avec le passage de l'analogique au numérique dans les années 1990
- Commence alors l'ère de la « dématérialisation »

Un ensemble complexe de technologies

Infrastructure



Python stack



Data Visualization



Big data processing



Data storage / formats / querying





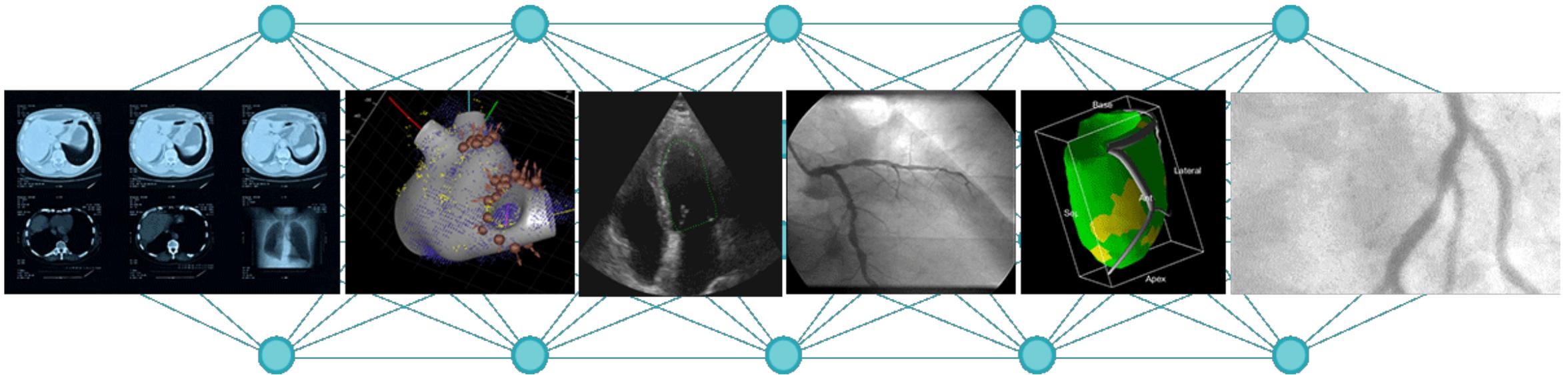
ÉTAT DES LIEUX

Qualité de vie Santé / Médecine

Des progrès en :

- Génomique
- Diagnostic sur image (cancers du sein, maladies de l'œil, etc.)
- Prévention, suivi des maladies chroniques
- Pharmacologie (mise au point de médicaments, vaccins)

DOSSIER | 13.09.2018
Comment le Big data
révolutionne la recherche
en santé
BIG DATA OMICS



Computer Vision in Cardiology, Global Leader in Computer Vision & Deep Learning (RSIP Vision)

Qualité de vie

Santé / Médecine



Hébergement de données de santé du Health Data Hub chez Microsoft : saisine du Ministre de la Santé pour potentiel manquement aux règles de la commande publique **Mars 2020**

HEALTH DATA HUB: LA CNIL DEMANDE L'ARRÊT DE L'HÉBERGEMENT DES DONNÉES DE SANTÉ PAR MICROSOFT **Octobre 2020**

Health Data Hub: Véran s'engage à retirer l'hébergement à Microsoft d'ici «deux ans»
Novembre 2020



VACCINATION: DOCTOLIB PASSE LE CAP DES 5 MILLIONS DE RENDEZ-VOUS **5 mars 2021**

Des syndicats de médecins contestent devant le Conseil d'Etat le partenariat entre l'Etat et Doctolib qui héberge les données chez Amazon **8 mars 2021**

Doctolib : le chiffrage des données incomplet ?

8 mars 2021

Campagne de vaccination : le Conseil d'Etat appelé à se prononcer sur le partenariat avec Doctolib **8 mars 2021**

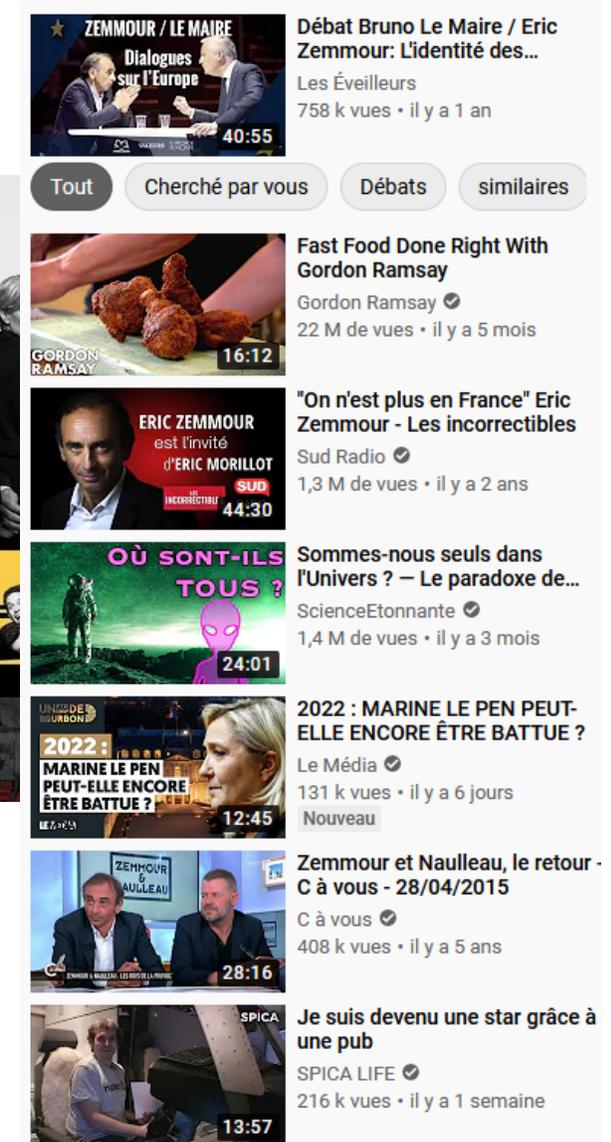
Qualité de vie

Algorithmes de recommandation

Normalisation de la culture, des mœurs, des idées et des imaginaires

Economie de l'attention :

- 70% du temps passé sur YouTube est dû aux algorithmes de recommandations
- 80% du contenu regardé sur Netflix est déterminé par les algorithmes



Qualité de vie

Des modes de vie
toujours plus incitatifs



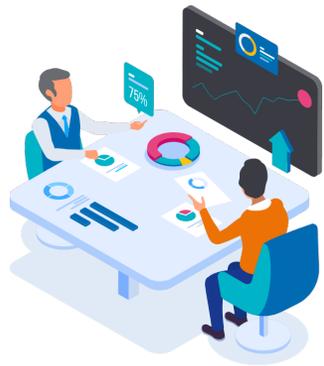
"Intelligence Artificielle"
Injonction
Coercition



Apprentissage profond
Incitation



Apprentissage automatique
Suggestions
Recommandations



Statistiques
Informatique décisionnelle
Aide à la décision

PROGRÈS

La complexification technologique de la société s'accompagne de :

- Structures de pouvoir toujours plus asymétriques
- Perte d'autonomie
- Désengagement politique

Environnement

Le mythe de la dématérialisation

- Sur les cinq dernières années, environ 1,4 milliards de smartphones ont été vendus dans le monde chaque année et un seul smartphone requiert environ 60 éléments différents (principalement des métaux rares ou critiques)
- La fabrication d'un ordinateur de 2kg mobilise environ 800kg de matières premières

Ressources nécessaires pour produire un ordinateur

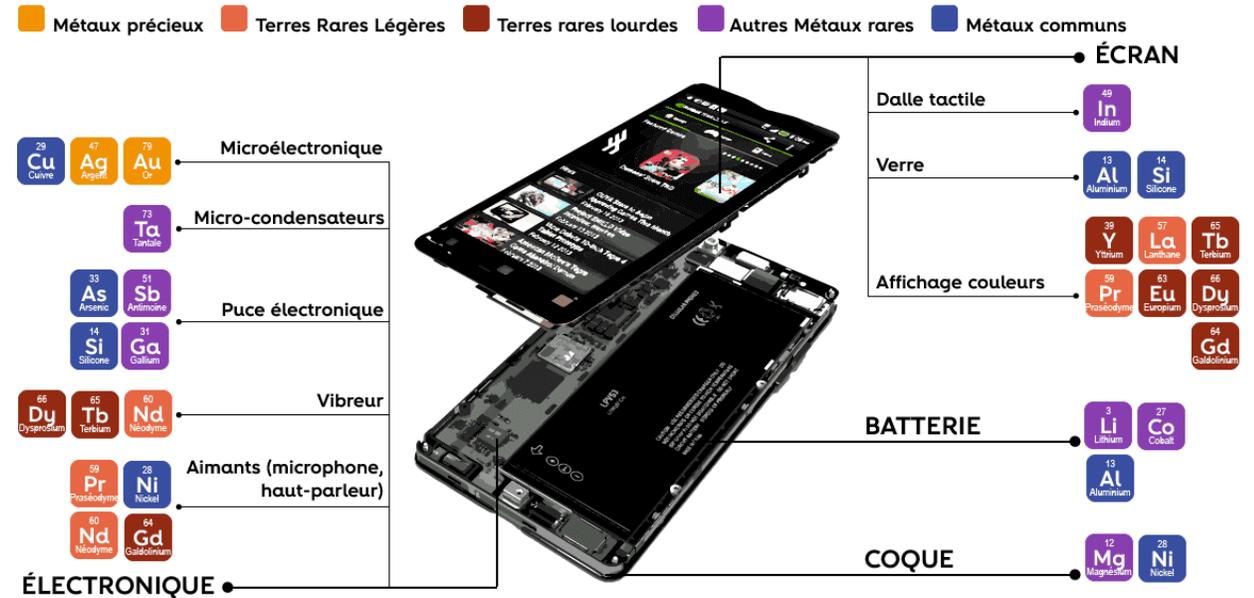


Source : Ademe

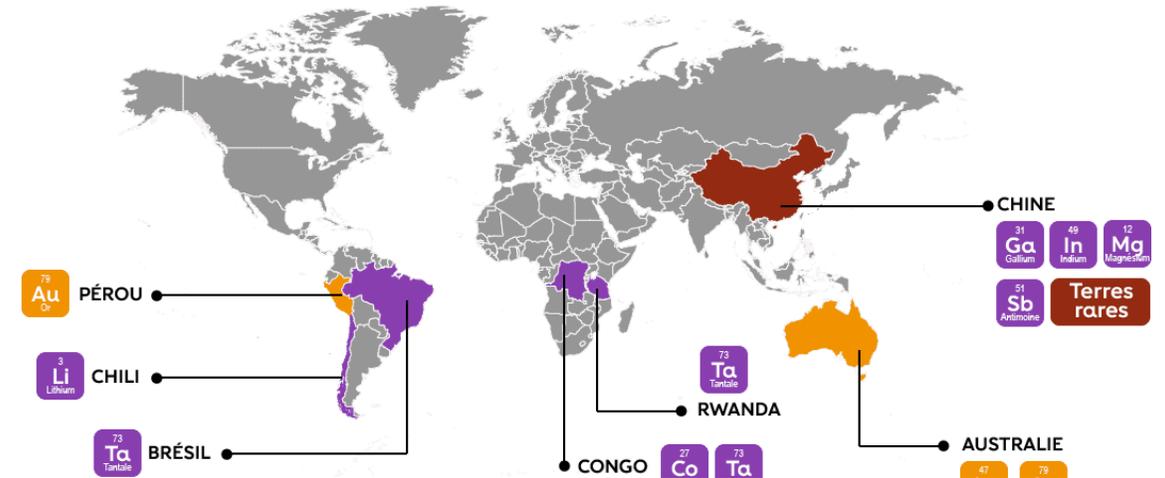
réalisé par Act 4 Earth

- Après le bâtiment, l'industrie technologique est le secteur le plus extracteur de sable
- Conséquences de cette surexploitation : recul des terres, disparitions de certaines îles, instabilité des infrastructures, danger pour l'agriculture... changement climatique

MÉTAUX CONTENUS DANS NOS SMARTPHONES



ORIGINE GÉOGRAPHIQUE DES COMPOSANTS



[Infographie BFM, 2018]

Environnement

Le mythe de la dématérialisation

- Nous avons multiplié les usages de ces métaux dans deux domaines qui sont les piliers essentiels de la *transition* énergétique:
 - **Les technologies supposément « vertes »;**
 - **Le numérique.**
- Car c'est de la convergence des *green tech* et de l'informatique que devrait naître un monde meilleur...
- Les premières (éoliennes, panneaux solaires, véhicules électriques), grâce aux métaux rares dont elle sont truffées, produisent une énergie décarbonée qui va transiter par des réseaux d'électricité dits « ultra-performants » qui permettent des économies d'énergies.
- Or ceux-ci sont pilotés par des technologies numériques, elles aussi farcies de tels métaux.

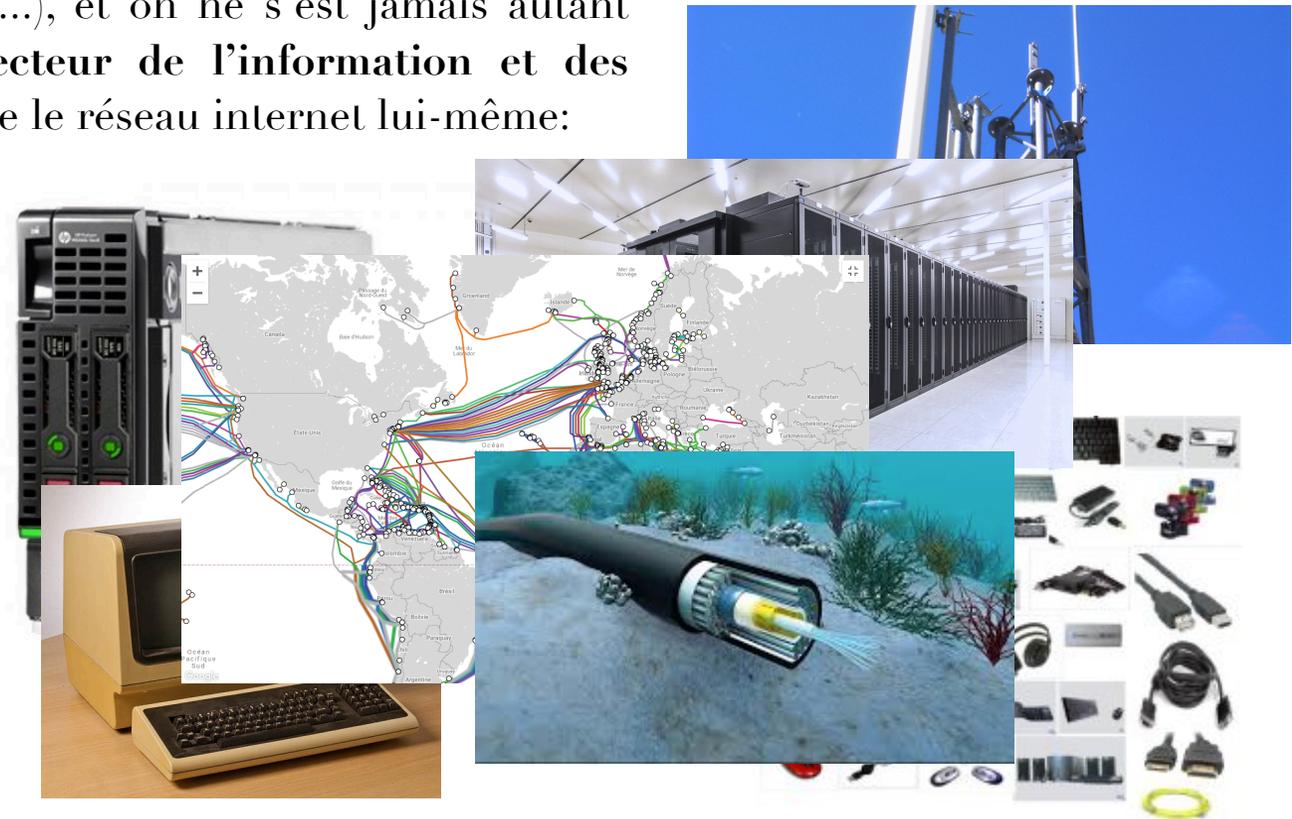
Ressource	Utilisations
Antimoine	Retardants du feu (additifs dans les plastiques), catalyse du polyéthylène
Baryte	Boues de forage pétroliers et gaziers, industrie du verre, radioprotection, santé, métallurgie, pyrotechnie
Béryllium	Télécoms et électronique, industrie aérospatiale, nucléaire civil et militaire
Cobalt	Portables, ordinateurs, véhicules hybrides, aimants
Germanium	Photovoltaïque, fibres optiques, catalyse, optique infrarouge
Indium	Puces électroniques, écrans LCD
Graphite naturel	Véhicules électriques, aérospatiale, industrie nucléaire
Silicium métal	Circuits intégrés, panneaux photovoltaïques, isolateurs électriques
Terres rares (voir annexe pour détails)	Aimants permanents, voiture électrique, éoliennes, TGV, scanners médicaux, lasers, transmission de données par fibre optique ...

Principales utilisations industrielles de quelques minerais rares
[Source du tableau: Guillaume Pitron]

Environnement

Le mythe de la dématérialisation

- Outre un pervers *effet rebond* on n'a presque jamais autant consommé de papier (entre 2000 et 2010 la consommation n'a baissé que de 1,3%...), et on ne s'est jamais autant déplacé pour des raisons professionnelles le secteur de l'information et des télécommunications n'a bien sûr rien de virtuel, comme le réseau internet lui-même:
 - Serveurs ;
 - Antennes-relais ;
 - Terminaux ;
 - Accessoires ;
 - Câbles transocéaniques de fibres optiques ...
- Les câbles sous-marins sont parmi les infrastructures les plus coûteuses du globe.
- Environ 1,2 millions de kilomètres de câbles de télécommunications sont déployés au fond des océans.

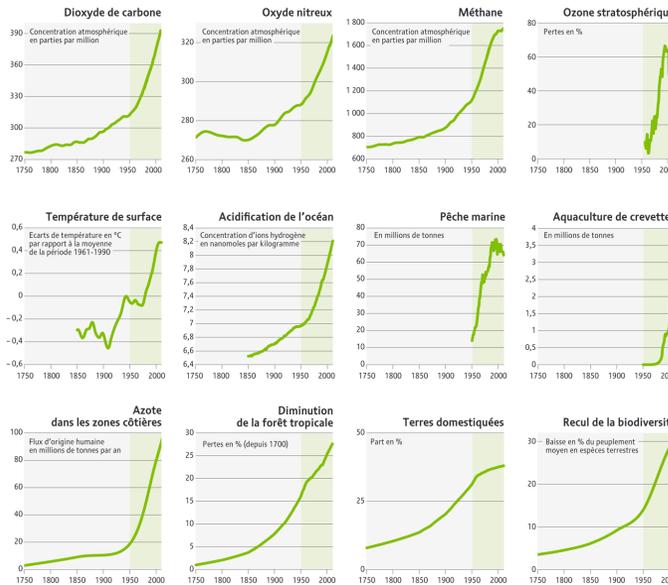


Environnement

La « grande accélération »

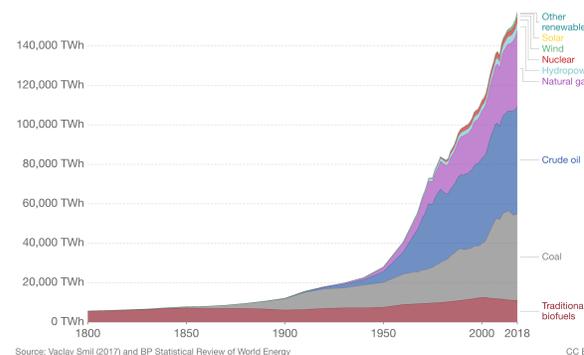
- Représente un tableau de bord de l'Anthropocène avec l'évolution de 24 paramètres du système Terre depuis 1750.
- Pour les 9 les plus significatifs, des scientifiques se sont intéressés aux points de basculement concernant la biodiversité, la pollution de l'air et de l'atmosphère, la perturbation des cycles biogéochimiques ou l'anthropisation des terres.
- Une limite à ne pas franchir pour ces 9 grands paramètres a été évaluée. Mais pour 4 d'entre eux, le seuil de danger de basculement brutal du système Terre vers des états catastrophiques est d'ores et déjà approchée ou dépassée : cycles de l'azote, émissions de GES, extinction de la biodiversité, cycle du phosphore

Evolution du système Terre

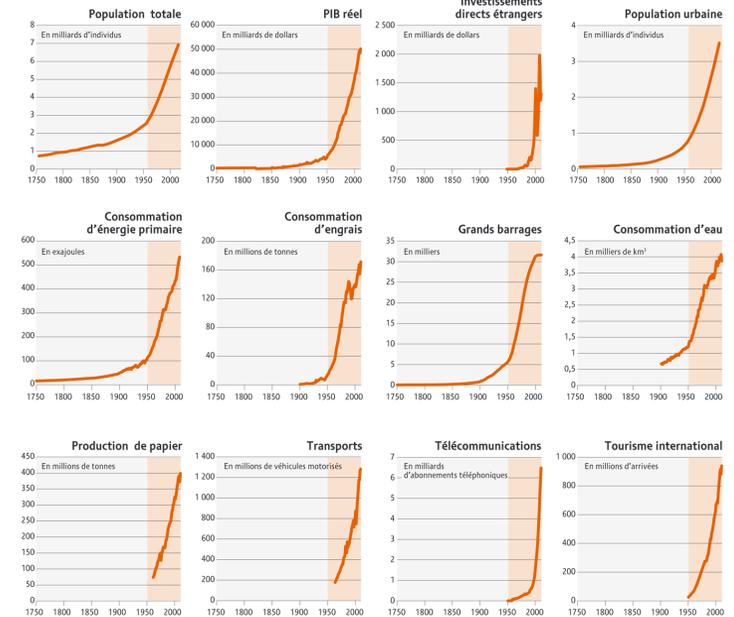


Global primary energy consumption

Global primary energy consumption, measured in terawatt-hours (TWh) per year. Here 'other renewables' are renewable technologies not including solar, wind, hydropower and traditional biofuels.

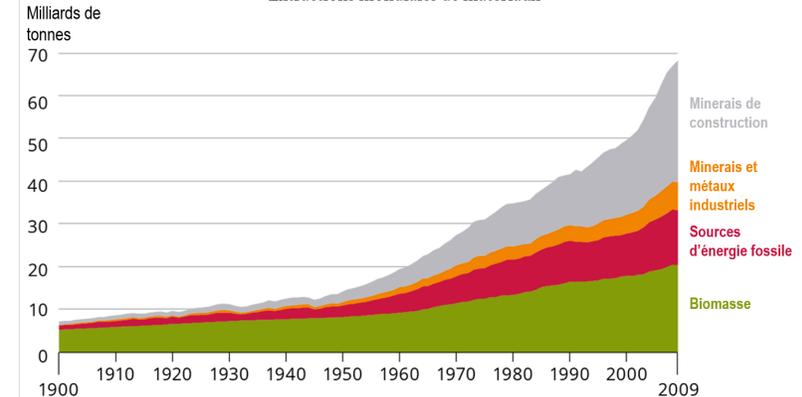


Développement socio-économique



Source : Will Steffen, Wendy Broadgate, Lisa Deutsch, Owen Gaffney et Cornelia Ludwig, « The trajectory of the Anthropocene : the Great Acceleration », *The Anthropocene Review*, 2015 (données : International Geosphere-Biosphere Programme et Stockholm Resilience Centre).

Extractions mondiales de matériaux



Environnement

Consommation énergétique et empreinte carbone du numérique

- La phase de production des équipements occupe une part très significative, **environ 45% en 2020**, dans l'empreinte énergétique totale du Numérique, ainsi que dans les émissions de GES qui en découlent.
- Un utilisateur de smartphone (s'il garde son appareil 2 ans) verra ainsi la **consommation énergétique totale induite au cours du cycle de vie de cet équipement se réaliser à plus de 90% avant même son achat***. Ce poids de la phase de production dans l'impact énergétique est de l'ordre de 60% pour une télévision connectée et >80% pour un ordinateur portable.
- Une bonne partie des enjeux environnementaux du Numérique n'est donc pas liée à son usage, mais en grande partie au volume de matériel produit, à son processus de production, et à sa durée de vie.

Common carbon footprint benchmarks

in lbs of CO2 equivalent

Roundtrip flight b/w NY and SF (1 passenger) | 1,984

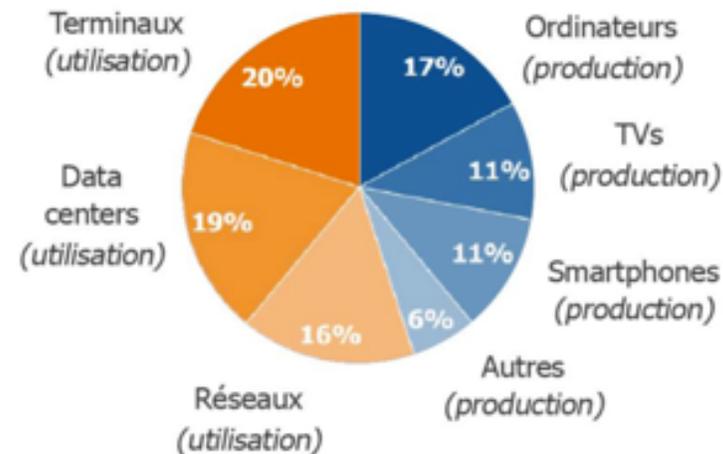
Human life (avg. 1 year) | 11,023

American life (avg. 1 year) | 36,156

US car including fuel (avg. 1 lifetime) | 126,000

Transformer (213M parameters) w/ neural architecture search | 626,155

Chart: MIT Technology Review • Source: Strubell et al. • Created with Datawrapper



Distribution de la consommation énergétique du Numérique par poste (2017)

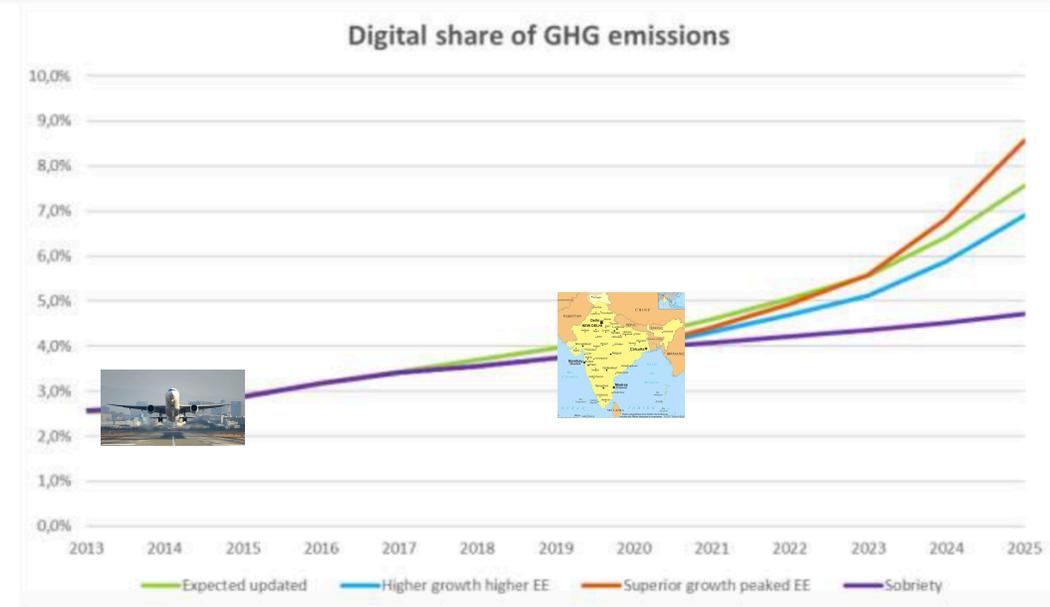
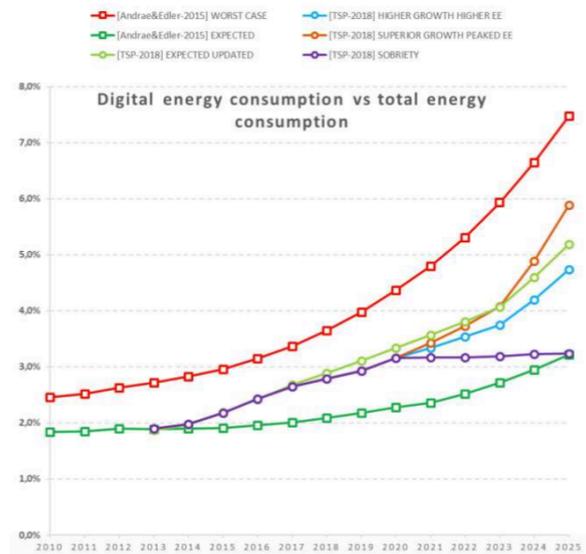
[Source : [Lean ICT Materials]]

* Nous parlons bien ici de la consommation électrique propre du smartphone : si l'on prend en compte également le surcroît de consommation électrique du réseau occasionné par l'utilisation du smartphone, la proportion est plutôt de l'ordre de 50%. Voir (Ercan, 2013).

Environnement

Consommation énergétique et empreinte carbone du numérique

- La part de **consommation énergétique** du Numérique passe de 1,9% en 2013 à 2,7% en 2017, et 3,3% en 2020
- Compte tenu du mix électrique mondial, la part **d'émissions de GES** attribuable au Numérique passe ainsi de 2,5% en 2013 à 4% en 2020 (2,1 Gt)
- En comparaison, la part d'émissions de GES des véhicules légers est d'environ 8% en 2018, et celle du transport aérien civil d'environ 2% en 2018. En 2020, le Numérique émettait autant de CO2 que l'Inde (1,3 milliards d'habitants) en 2015.



- Non seulement ces dispositifs sont coûteux en termes d'extraction, d'acheminement et de transformation de matières premières, mais ils sont responsables d'au moins 10% de la consommation électrique mondiale, avec une tendance à la hausse.
- En effet, la croissance est fulgurante : **90 % des données actuelles ont moins de 2 ans d'existence** (CNRS, 2018). Si cette cadence était ou pouvait être maintenue, en 2037, les outils et infrastructures informatiques consommeraient plus d'électricité que le monde ne serait en mesure d'en produire (Usine Nouvelle, 2018).

Environnement

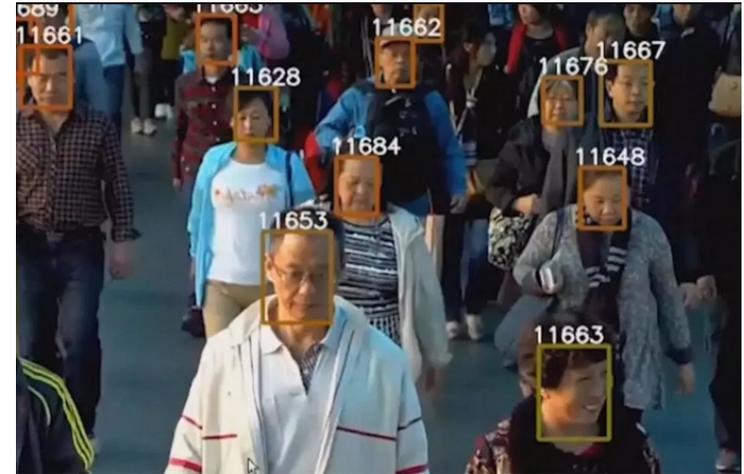
Qu'en pensez-vous ?

- À votre avis, le big data a-t-il un rôle à jouer dans la crise écologique ?
- Si oui, lequel ?

- Améliorer les prévisions de consommation d'énergie
- Réduire les obstacles à l'adoption des véhicules électriques
- Découvrir de nouveaux matériaux
- Optimiser les chaînes d'approvisionnement
- Aider à rendre les bâtiments plus efficaces
- Rendre possible l'agriculture de précision à l'échelle
- Améliorer le suivi de la déforestation
- Pousser les consommateurs à changer leur façon de consommer

Démocraties et libertés

Surveillance



- Les géants du numérique ont désormais la mainmise sur les services informatiques, les infrastructures et les données des utilisateurs.
- Un pouvoir qui leur donne un poids considérable dans la politique des États occidentaux, dont les gouvernements n'hésitent pas à collaborer avec ces multinationales pour se maintenir.
- Le crédit social chinois, paroxysme de l'hyper-rationalisation rendue possible par le Big Data, témoigne des effets de centralisation du pouvoir et de son incompatibilité avec la démocratie.
- Voir sur ARTE : *Tous surveillés - 7 milliards de suspects*

Démocraties et libertés

Surveillance

Valenciennes - 2019

Installation de 240 caméras offertes par Huawei et équipées d'un système de reconnaissance faciale

Metz - 2016

Installation d'un dispositif de reconnaissance faciale dans un stade de foot

Paris - 2020

Détection du port du masque dans le RER C

Saint-Etienne - 2019

Un projet de surveillance sonore est retoqué par la CNIL

Nice - 2019

- Système de reconnaissance faciale israélien testé sur la foule du carnaval
- Détection des émotions des passagers de tramways
- Portiques de reco faciale dans deux lycées

Cannes - 2020

Détection du port du masque

Hérault - 2019

Détection de bâtis non déclarés par images satellite

Marseille - 2019

Caméras détectant des comportements suspects sur la voie publique

Loi Sécurité Globale validée au Sénat le 3 mars 2021 :

- Autorisation et banalisation de la surveillance par drone
- Reconnaissance faciale en direct (caméras drone et piéton)
- Mutualisation des moyens de surveillance publics/privés sur tout le territoire

Décret du 10 mars 2021

Autorisation au recours à la vidéo intelligente pour mesurer le taux de port de masque dans les transports

Décrets PASP du 8 novembre 2020

- Fichage massif de militants et d'opposants politiques (et de leur entourage)
- Surveillance automatisée sur les réseaux sociaux
- Recensement des opinions politiques et des données de santé
- Fichage des victimes
- Elargissement de la reconnaissance faciale

Start-up Nation

Démocraties et libertés

Les données du Grand Débat

Une opportunité unique d'analyser longuement de grands volumes d'expression démocratiques, en mettant à contribution notamment des ingénieurs (algorithmes de traitement du langage naturel), des linguistes, des sociologues, des historiens, etc. Mais :

- L'analyse des données du Grand Débat est confiée, **sans appel d'offre**, à un consortium de 3 entreprises dirigé par le cabinet Roland Berger.
- Une synthèse est rédigée et publiée, sans que les revendications du mouvement des GJ soient entendues
- Les cahiers de doléances recueillis par l'État ne sont **jamais publiés...**
- ... Sous prétexte d'un trop grand volume de données.



Synthèses des contributions en ligne

Questionnaires rapides et propositions partagées sur la plateforme





SUGGESTIONS

Déconstruire

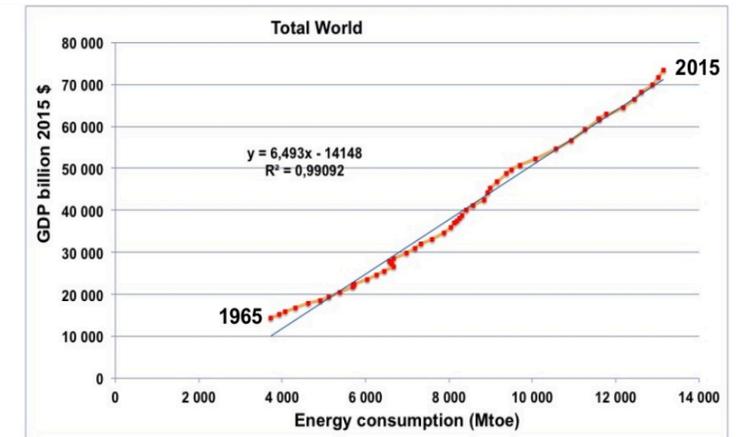
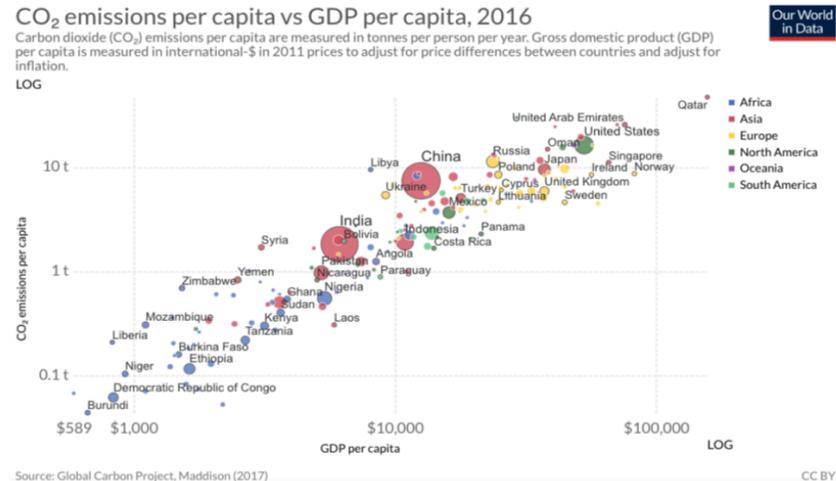
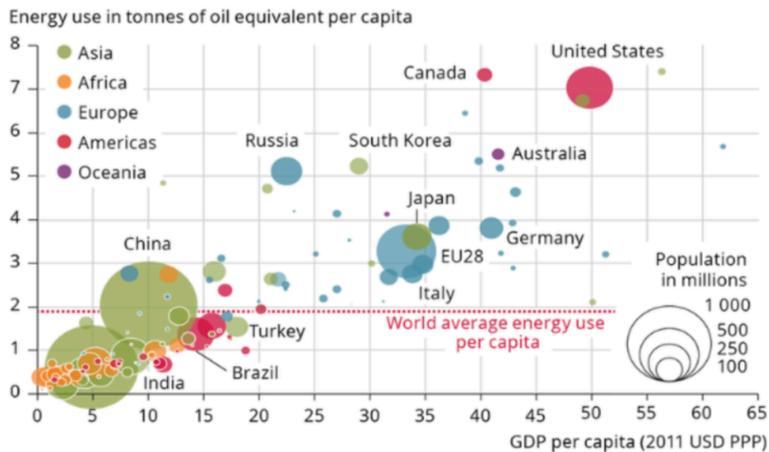
La croissance économique

- Dans sa thèse sur l'économie politique de la décroissance, T. Parrique synthétise une triple critique de la croissance (ch. 2, 3, 4)
 1. **Limites biophysiques** : la croissance est *insoutenable* parce qu'on ne peut pas découpler PIB et pressions environnementales
 2. **Limites sociales** : la croissance est *insupportable* car, passé certains seuils, elle peut détruire du lien social
 3. **Limites politiques** : la croissance est *inutile* car elle ne garantit pas l'emploi, ne réduit ni la pauvreté ni les inégalités, et n'augmente pas le bien être

Déconstruire

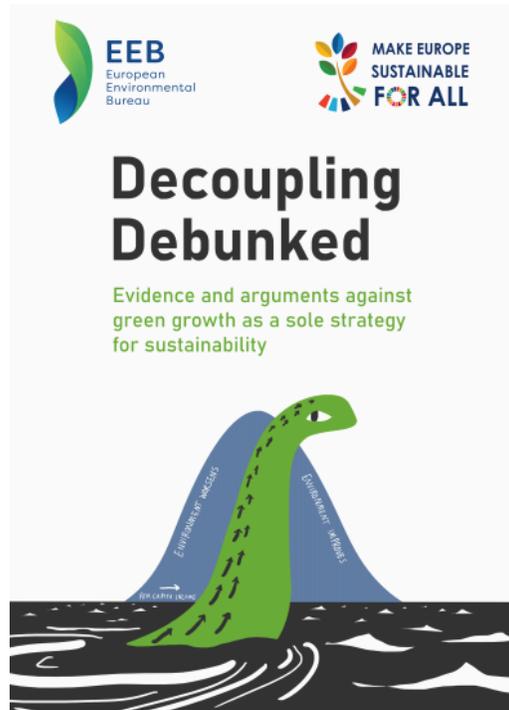
Le découplage

- Quel est le lien entre PIB et pressions environnementales (consommation d'énergies, de ressources, GES, pollutions, destructions d'habitats naturels, d'espèces etc.) ?



Energie consommée (en abscisse) et PIB en dollars constants (ordonnée) pour le monde, de 1965 à 2015. Données World Bank pour le PIB et BP stat pour l'énergie

- Plus qu'une (très forte) corrélation, une relation de causalité (Giraud, 2014)



ENVIRONMENTAL RESEARCH LETTERS

TOPICAL REVIEW

A systematic review of the evidence on decoupling of GDP, resource use and GHG emissions, part I: bibliometric and conceptual mapping

Dominik Wiedenhofer¹ , Doris Virág¹ , Gerald Kalt¹ , Barbara Plank¹ , Jan Streeck¹ ,
Melanie Pichler¹ , Andreas Mayer¹ , Fridolin Krausmann¹ , Paul Brockway² , Anke Schaffartzik^{1,7} 

[+ Show full author list](#)

Published 11 June 2020 • © 2020 The Author(s). Published by IOP Publishing Ltd

[Environmental Research Letters](#), Volume 15, Number 6

Citation Dominik Wiedenhofer et al 2020 *Environ. Res. Lett.* 15 063002

Déconstruire

Le découplage

- Le « développement durable » est une (longue) quête de « découplage » entre croissance énergétique (fossile, et donc 1ère cause du réchauffement climatique) ou matérielle (extractivisme) et croissance économique. On parle aussi de « croissance verte ».
- Or ce « découplage » ne vient pas, n'a jamais été observé, et fait l'objet d'une remise en question empirique grandissante.
- Ces connaissances sont rassemblées dans le rapport « Decoupling Debunked » de l'EEB (2019).
- En 2020, 16 chercheurs publient la première étude systématique de la totalité des papiers sur le découplage (11 500 papiers, 835 études empiriques).
- Résultat inchangé : les taux de découplage observés sont insignifiants.

Déconstruire

Les concepts obsolètes (en suggérer des nouveaux)

- D. Landivar et A. Monnin, enseignants-chercheurs en économie et humanités numériques à l'ESC Clermont, proposent de :
 - Ne plus utiliser le concept de « développement durable »
 - On lui préférera celui de transition anthropocénique : versant réaliste de la *transition écologique*, explicitant ce qui doit être dépassé
 - Se débarrasser de la « RSE (Responsabilité Sociale de l'Entreprise) »
 - Pour aller vers la notion de Responsabilité Héritée des Entreprises (RHE) : s'apparentant plus à un réseau d'attachement qui nous oblige, *malgré nous*, à vivre avec le capitalisme (ou ce qu'il en reste)

Déconstruire

Les grands mensonges verts

Twelve years of carbon neutrality



In 2007, Google committed to being carbon neutral, and we've met this goal every year since then. We reach carbon neutrality via three steps. First, we work to reduce our total energy consumption by pursuing aggressive energy-efficiency initiatives. Second, we match 100% of the electricity consumption of our operations with purchases of renewable energy. Third, we buy carbon offsets for any remaining emissions we haven't yet eliminated.⁴⁰



Figure 7

GHG EMISSIONS WITHOUT RENEWABLE ENERGY PURCHASES

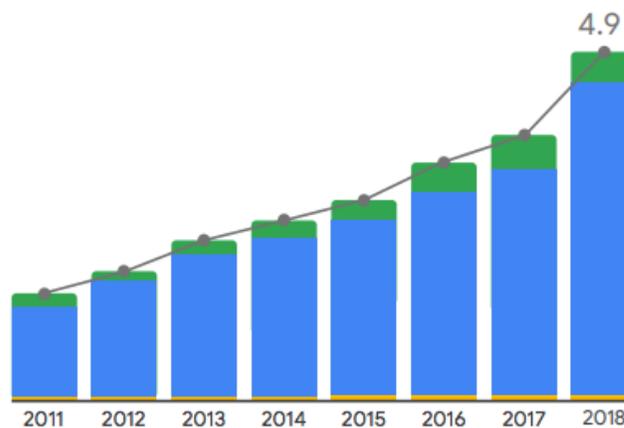
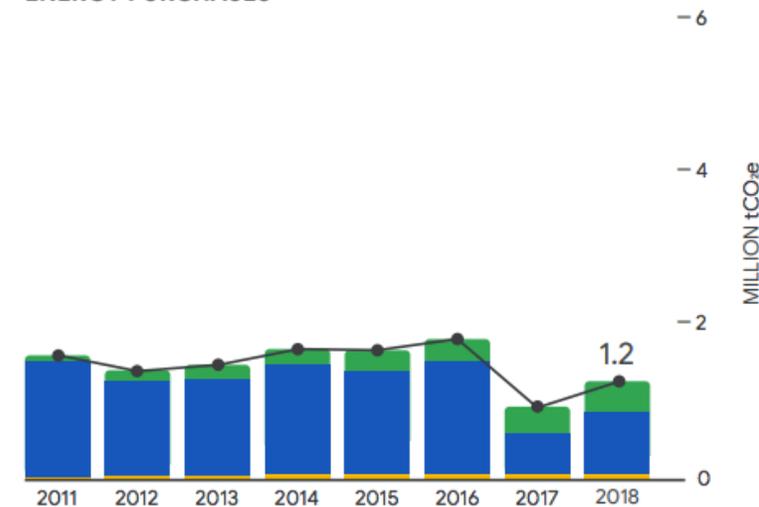


Figure 8

GHG EMISSIONS WITH RENEWABLE ENERGY PURCHASES



■ Scope 1 ■ Scope 2 (location) ■ Scope 2 (market) ■ Scope 3 (business travel and commuting)³⁸
● Total location-based GHG emissions ● Total market-based GHG emissions

Déconstruire

Les grands mensonges verts

nature climate change

Explore Content ▾ Journal Information ▾ Publish With Us ▾

nature > nature climate change > correspondence > article

Correspondence | Published: 13 July 2020

Hiding greenhouse gas emissions in the cloud

David Mytton 

Nature Climate Change 10, 701(2020) | Cite this article

4884 Accesses | 56 Altmetric | Metrics



- Si elles sont délocalisées dans le cloud, les émissions IT précédemment comptées en :
 - **Scope 1** (*sources possédées ou contrôlées par une entreprise*) et
 - **Scope 2** (*énergie achetée*) passent au
 - **Scope 3** (*émissions indirectes qui interviennent dans la chaîne de valeur*).

- Quand une organisation achète son propre équipement et signe des contrats pour l'électricité d'une installation de colocation, ses émissions IT relèvent du Scope 1 ou 2. Si la charge de travail IT est déplacée vers le cloud, l'organisation ne génère plus directement d'émissions du Scope 1 et n'achète plus d'énergie du Scope 2. Par conséquent, toutes ces émissions passent dans le Scope 3.
- Les émissions du Scope 3 ne sont déclarées que sur une base volontaire, et les émissions sont regroupées dans les rapports sur les émissions globales par les grands fournisseurs de services cloud, ce qui peut avoir pour conséquence de **masquer les émissions d'une entreprise individuelle lorsqu'elles sont déplacées vers le cloud.**
- C'est un problème car lorsque les règles nationales de déclaration exigent la divulgation de statistiques sur les émissions, seuls les **Scope 1 et 2 sont obligatoires. [...] Le problème, c'est l'opacité.**

OSINT Landscape v.1 February 2018

Open Source Intelligence (/OSINV – Open Source Investigation)

COVERT SHORES *bellingcat*
www.hisutton.com

Social Media Platforms

Facebook Weibo Twitter Qzone Instagram Odnoklassniki LinkedIn VK Snapchat YouTube Periscope

stalkscan Facebook Search Tool Graph Search Generator f FBDOWN f Signal peoplefindThor

Tweetbeaver twXplorer Twitter List Copy TweetDeck Download Twitter Videos

WEBSTA Instagram Downloader picodash

socilab

PHOTO MAP

saverfrom.net Youtube DataViewer frame by frame storyful. Geo Search Tool

Scopedown

Snap Map

Dataminr INTELTECHNIQUES Echosec War Wire

Sharing & Publishing

flickr Pinterest Google+

Blogging, Forums & other communities

STRAVA tumblr. LIVEJOURNAL classmates Medium

WORDPRESS.ORG Blogger WIX.com

ProBoards SQUARESPACE Joomla! ghost weebly

Internet Search

Google Yandex Bing

waybackMachine DuckDuckGo NAVER

Baidu 百度 搜狗搜索 goo

Рамблер/ kakao

YAHOO! MAKTOOB مکتوب archive.today

PimEyes

Geospatial Data

GeoNames Free GIS Data

SECRETS OF THE WEST OpenRailwayMap MAPS.ME

DualMaps wikimapia

Google Maps Bing maps Apple Maps here Yandex Maps

Mapillary

Maritime Movements

MarineTraffic IHS Markit AISLive Lloyd's List Intelligence SHIPSPOTTING.COM

Vessel Finder Shipfinder AISHub AIS Data Exchange shipfinder

OpenSeaMap SHIPAIS ShippingExplorer BoatNerd AisDECODER

Aviation Movements

AirNav.RadarBox LiveATC.net ADS-B Exchange GVA Dictator Alert FlightAware PLANESPOTTERS.NET

COAA

Radio

RadioReference Broadcastify Radio Garden SDR.HU Global Tuners ProScan MilScanners

Commercial Registries

opencorporates infobel

ICU OFFSHORE LEAKS DATABASE Investigative Dashboard Search EUROPEAN JUSTICE

Webcams

pentopia Free Live Webcams Insecam SHODAN EarthCam Webcams.travel PICTIMO wetter.com lookr wisuki

Image / Vid / Doc Forensics

GET-METADATA Jeffrey's Image Metadata metapicz FotoForensics Forensically Beta IRFANVIEW hatlord / Spiderpig exifdata ExifTool izitru InVID

Satellite Imagery

Google Earth Descartes Labs TERRA server AIRBUS GeoStore Zoom Earth NEW SATELLITE IMAGES EVERY DAY planet unitar Radiant.Earth

NOAA EARTHDATA esa Earth Online USGS EarthExplorer esa opernicus DigitalGlobe SENTINEL

This landscape shows data sources (mostly platforms, tools or apps) that provide publicly available data which may be of use in OSINT. Some tools may charge for data access. It is intended to be extensive, but not exhaustive, and may be updated periodically.

Authors:
H I Sutton, (@CovertShores) Covert Shores and Jane's contributor,
Aliaume Leroy, (@Yalri) Bellingcat & BBC,
Tony Roper, (@Topo_MSS27), planesandstuff, Jane's contributor

Quelques notions d'OSINT

Qu'est-ce que l'OSINT ?



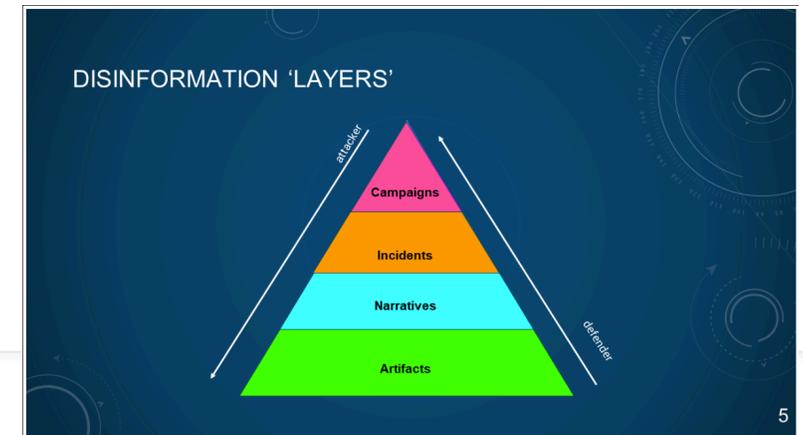
- Ensemble de disciplines visant à collecter et analyser des informations extraites de sources librement accessibles (sites Web, comptes sur les réseaux ou médias sociaux, imageries satellitaires, journaux papier, etc.).
- Plus qu'un assemblage d'outils et de méthodes, il est d'usage de parler d'un « état d'esprit », tourné vers l'investigation, fait de rigueur et d'éthique.
- Par nature transverse, l'OSINT peut être utilisé par différents acteurs dans de multiples domaines : journalisme, justice, sécurité informatique, intelligence économique, à titre individuel.

Quelques notions d'OSINT

#Ex 1 : « La détection de la désinformation sans coder » (OpenFacto, 2021)

Strategic Planning	Objective Planning	Develop People	Develop Networks	Microtargeting	Develop Content	Channel Selection	Pump Priming	Exposure	Go Physical	Persistence
4 items	2 items	3 items	6 items	3 items	10 items	10 items	8 items	10 items	2 items	3 items
5Ds (dismiss, distort, distract, dismay, divide)	Center of Gravity Analysis	Create fake experts	Create fake websites	Clickbait	Adapt existing narratives	Backstop personas	Bait legitimate influencers	Cheerleading domestic social media ops	Organise remote rallies and events	Continue to amplify
Competing Narratives	Create Master Narratives	Create fake or imposter news sites	Create funding campaigns	Paid targeted ads	Conspiracy narratives	Facebook	Demand unsurmountable proof	Cow online opinion leaders	Sell merchandising	Legacy web content
Facilitate State Propaganda		Create fake Social Media Profiles / Pages / Groups	Create hashtag	Promote online funding	Create competing narratives	Instagram	Deny involvement	Dedicated channels disseminate information pollution		Play the long game
Leverage Existing Narratives			Cultivate ignorant agents		Create fake research	LinkedIn	Kernel of Truth	Fabricate social media comment		
			Hijack legitimate account		Create fake videos and images	Manipulate online polls	Search Engine Optimization	Flooding		
			Use concealment		Distort facts	Pinterest	Seed distortions	Muzzle social media as a political force		
					Generate information pollution	Reddit	Use fake experts	Tertiary sites amplify news		
					Leak altered documents	Twitter	Use SMS/ WhatsApp/ Chat apps	Twitter bots amplify		
					Memes	WhatsApp		Twitter trolls amplify and manipulate		
					Trial content	YouTube		Use hashtag		

- Traiter ces campagnes comme des attaques informatiques (cf. COGSEC Collaborative Cognitive Security) et y appliquer la même analyse pour définir des contre-mesures
- S'inspirer d'un cadre de modélisation des procédés d'intrusion dans un système informatique : **AMITT Framework** (Adversarial Misinformation and Influence Tactics and Techniques)



Quelques notions d'OSINT

#Ex 1 : « La détection de la désinformation sans coder » (OpenFacto, 2021)

Strategic Planning	Objective Planning	Develop People	Develop Networks	Microtargeting	Develop Content	Channel Selection	Pump Priming	Exposure	Go Physical	Persistence
4 items	2 items	3 items	6 items	3 items	10 items	10 items	8 items	10 items	2 items	3 items
SDs (dismiss, distort, distract, dismay, divide) Competing Narratives Facilitate State Propaganda Leverage Existing Narratives	Center of Gravity Analysis Create Master Narratives	Create fake experts Create fake or imposter news sites Create fake Social Media Profiles / Pages / Groups	Create fake websites Create funding campaigns Create hashtag Cultivate ignorant agents Hijack legitimate account Use concealment	Clickbait Paid targeted ads Promote online funding	Adapt existing narratives Conspiracy narratives Create competing narratives Create fake research Create fake videos and images Distort facts Generate information pollution	Backstop personas Facebook Instagram LinkedIn Manipulate online polls Pinterest Reddit Twitter WhatsApp YouTube	Bait legitimate influencers Demand unsurmountable proof Deny involvement Kernel of Truth Search Engine Optimization Seed distortions Use fake experts Use SMS/ WhatsApp/ Chat apps	Cheerleading domestic social media ops Cow online opinion leaders Dedicated channels disseminate information pollution Fabricate social media comment Flooding Muzzle social media as a political force Tertiary sites amplify news Twitter bots amplify Twitter trolls amplify and manipulate Use hashtag	Organise remote rallies and events Sell merchandising	Continue to amplify Legacy web content Play the long game

Cette personne existe ? :
 - Botsentinel (donner un score à un compte pour identifier les bots)
 - Accountanalysis (analyse de l'activité du compte en question)

Partie visible et extrême de l'iceberg :
 lien entre évènement physique et origine numérique

Défi du plus grand nombre :
 - GetdayTrend (tendances virales),
 - Hoaxy (visualiser communautés et identifier bots potentiels)

Le temps de l'analyse :
 Utiliser sa tête pour articuler les résultats, essayer de comprendre le grand dessein d'une campagne de désinformation et son organisation
 Limites: temps & capacité de traitement / représentation de données massives

Faire l'état des lieux de la propagation de la campagne :
 CrowdTangle (proposé extension Chrome pour suivre propagation d'une url sur les réseaux appartenant à FB)

Ex #1 : « La détection de la désinformation sans coder », OpenFacto

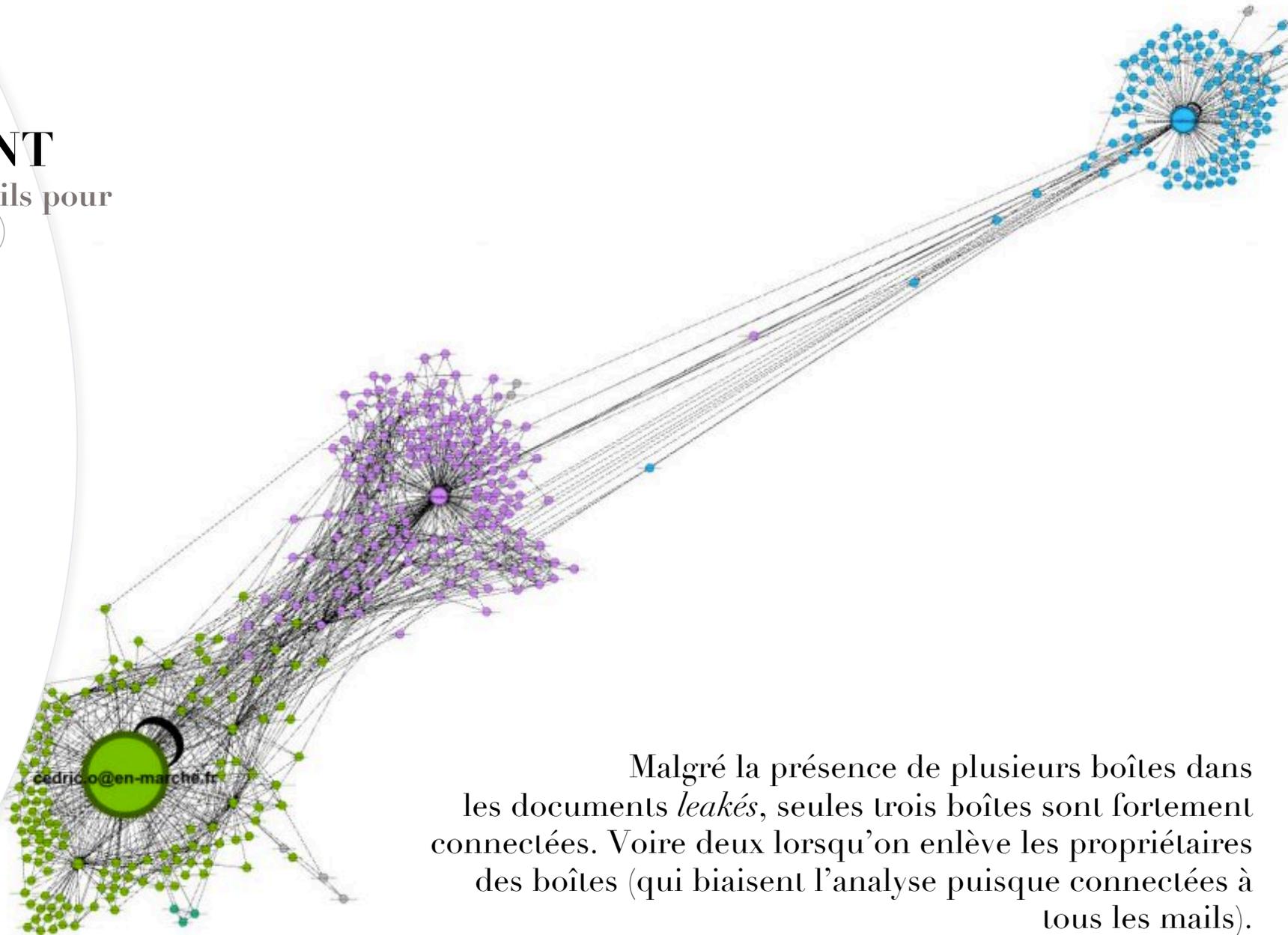
The screenshot displays the OpenFacto interface for analyzing a Twitter account. Key elements include:

- Twitter Trends Worldwide:** A list of trending topics for March 10, 2021, such as Messi, Dembele, Navas, Lenglet, and #PSGFCB.
- Account Analysis:** A section for 'accountanalysis' showing tweet volume by date, day of week, and tweet type. A 'Rating: Normal' is displayed with a 0% score.
- Network Graphs:** A 'CrowdTangle' style network graph showing interactions between accounts.
- Facebook Interactions:** A section showing interactions on Facebook, including likes and shares.
- IP Referrals:** A list of IP addresses and their associated accounts.

Quelques notions d'OSINT

#Ex 2 : « MacronLeaks : quelques outils pour explorer des mails » (Alphoenix, 2017)

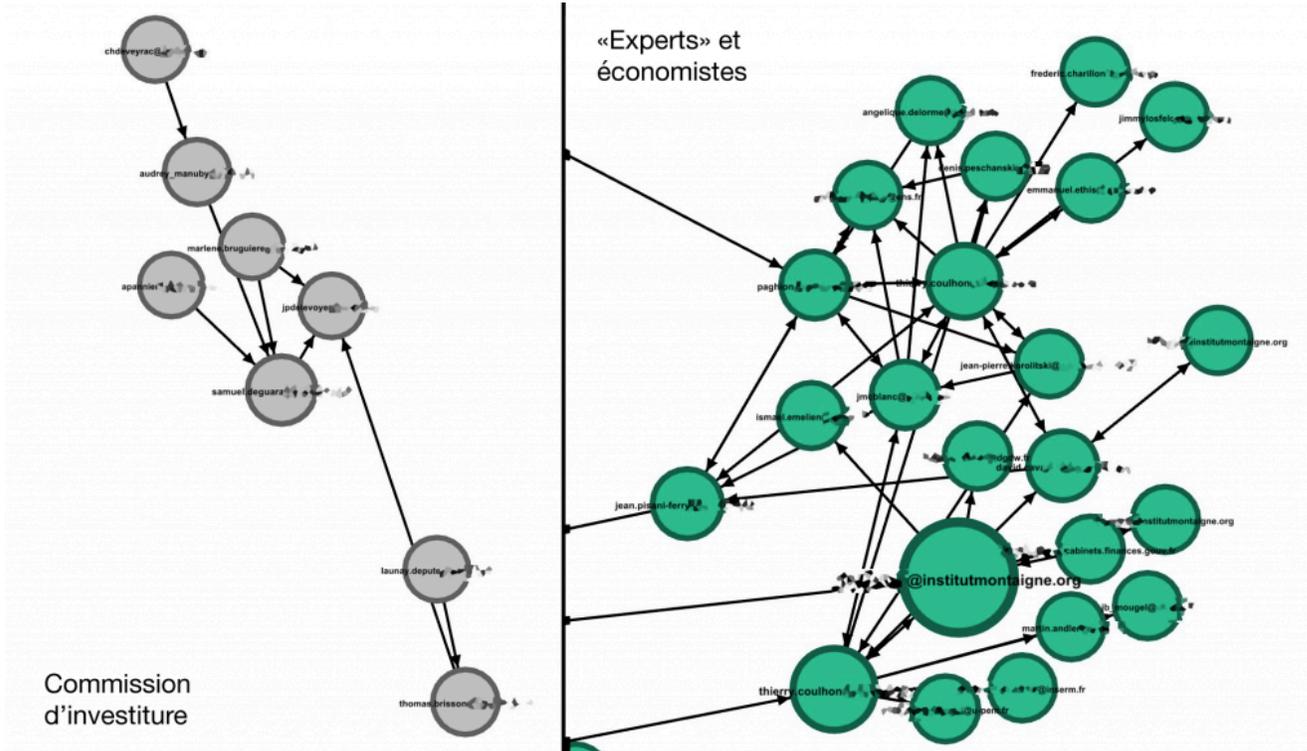
- Une autre analyse possible se fait grâce à **Gephi**.
- L'idée est de voir les réseaux qui se créent entre les différents expéditeurs et destinataires : plus deux personnes se parlent, plus proches elles seront sur le graphe. Apparition de petits groupes, expéditeurs les plus prolixes, rapprochement dans le temps...
- Le graphe ainsi réalisé compte 2 000 adresses mails et 12 000 relations.



Malgré la présence de plusieurs boîtes dans les documents *leakés*, seules trois boîtes sont fortement connectées. Voire deux lorsqu'on enlève les propriétaires des boîtes (qui biaisent l'analyse puisque connectées à tous les mails).

Quelques notions d'OSINT

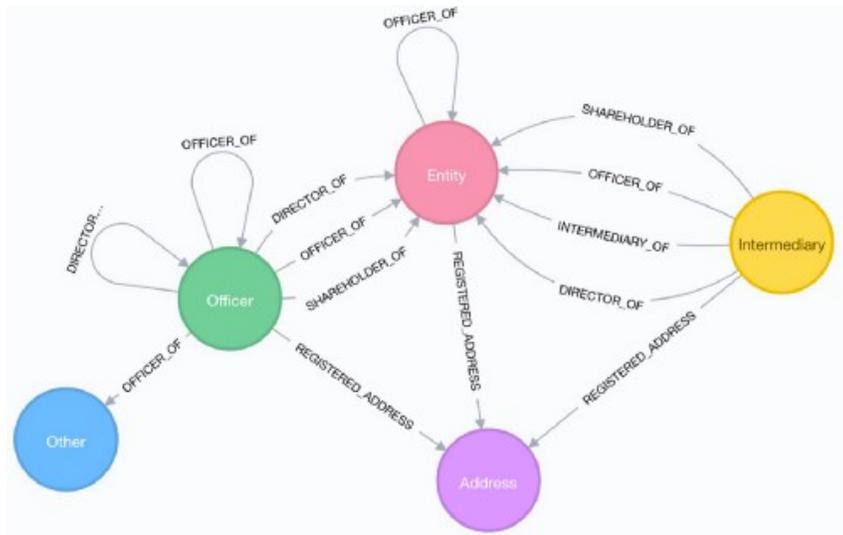
#Ex 2 : « MacronLeaks : quelques outils pour explorer des mails » (Alphoenix, 2017)



- Ce graphe permet néanmoins de faire apparaître quelques groupes créés pendant la campagne.
- On voit par exemple des mails échangés fréquemment entre les économistes proches de la campagne, ou alors ceux entre les membres de la commission d'investissement du parti.
- Gephi permet également de créer un graphe dynamique grâce à la date, permettant de voir le réseau se créer.
- D'autres outils permettent de générer une liste des termes revenant le plus, d'identifier des entités et de trouver rapidement par exemple des factures...

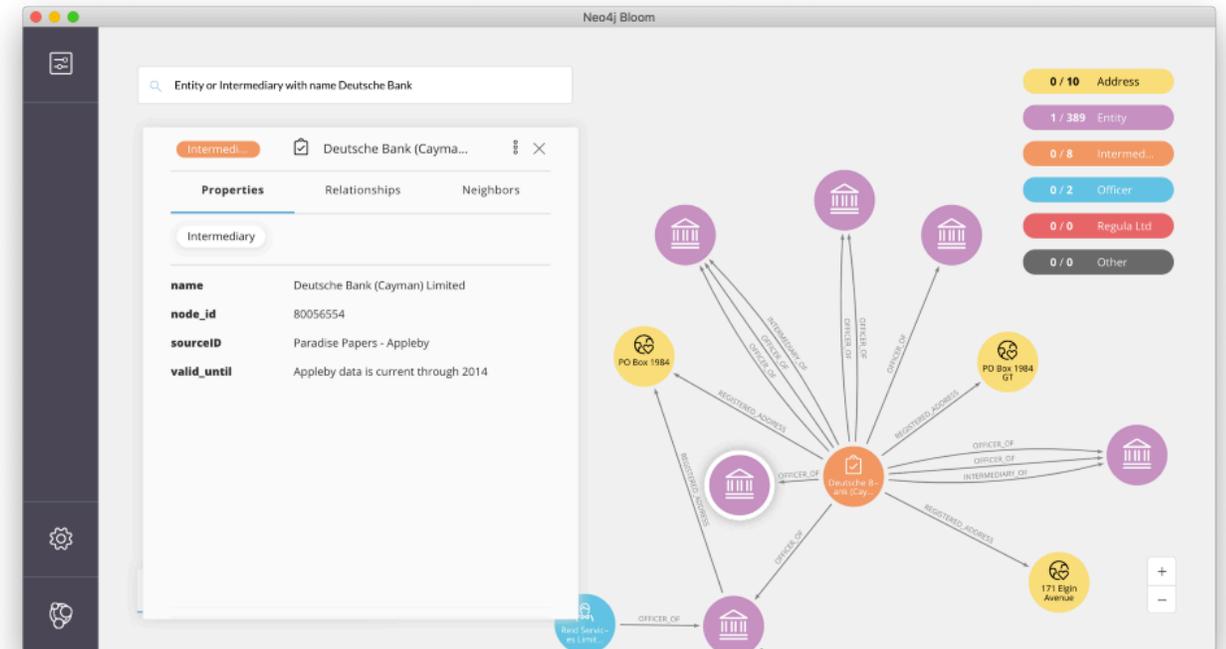
Quelques notions d'OSINT

#Ex 3 : « Graph Visualization of Panama Papers Data In Neo4j » (W. Lyon, 2018)



The Panama Papers / Offshore Leaks Neo4j database data model.
The database contains offshore legal entities and connections to them.

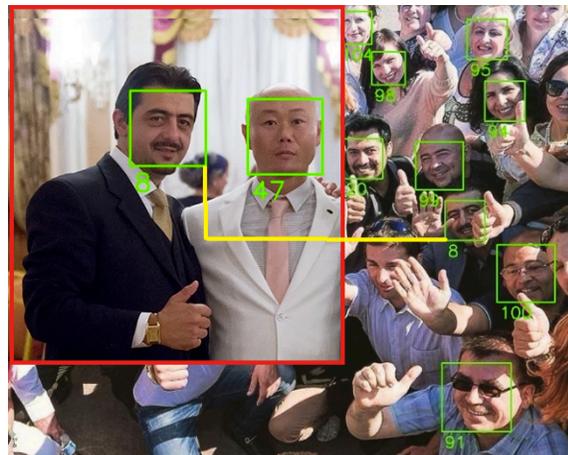
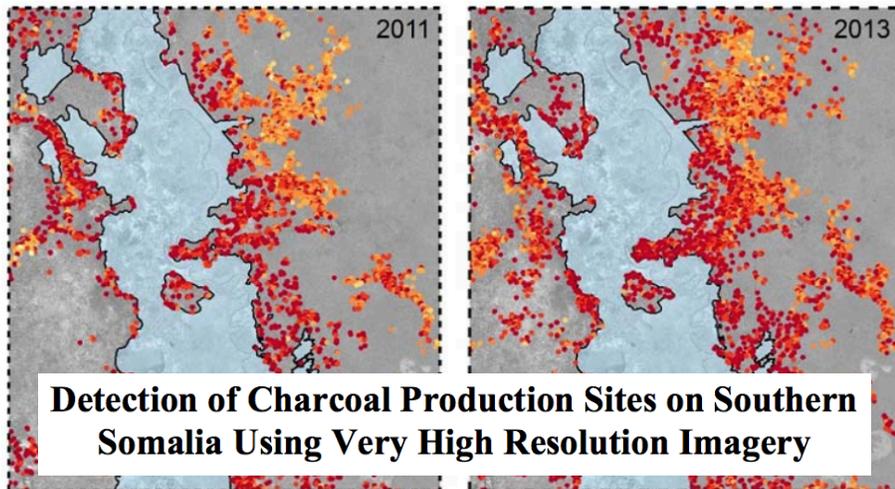
Inspecting a specific Deutsche Bank offshore legal entity.



- Nous inspectons ici l'une des correspondances de la Deutsche Bank, « Deutsche Bank (Cayman) Limited » et pouvons voir que cette entité sert de dirigeante à plusieurs autres entités juridiques offshore, et qu'elle a également servi « d'intermédiaire » pour plusieurs entités offshore.

Quelques notions d'OSINT

D'autres exemples



Who-Where-WhomWith(WWWW): A Facial Recognition Tool for Image-based Data Gathering and Graph Analysis

Lorenzo Romani [Follow](#)
May 13, 2020 · 7 min read



Quelques notions d'OSINT

Qu'est-ce que l'OpSec ?

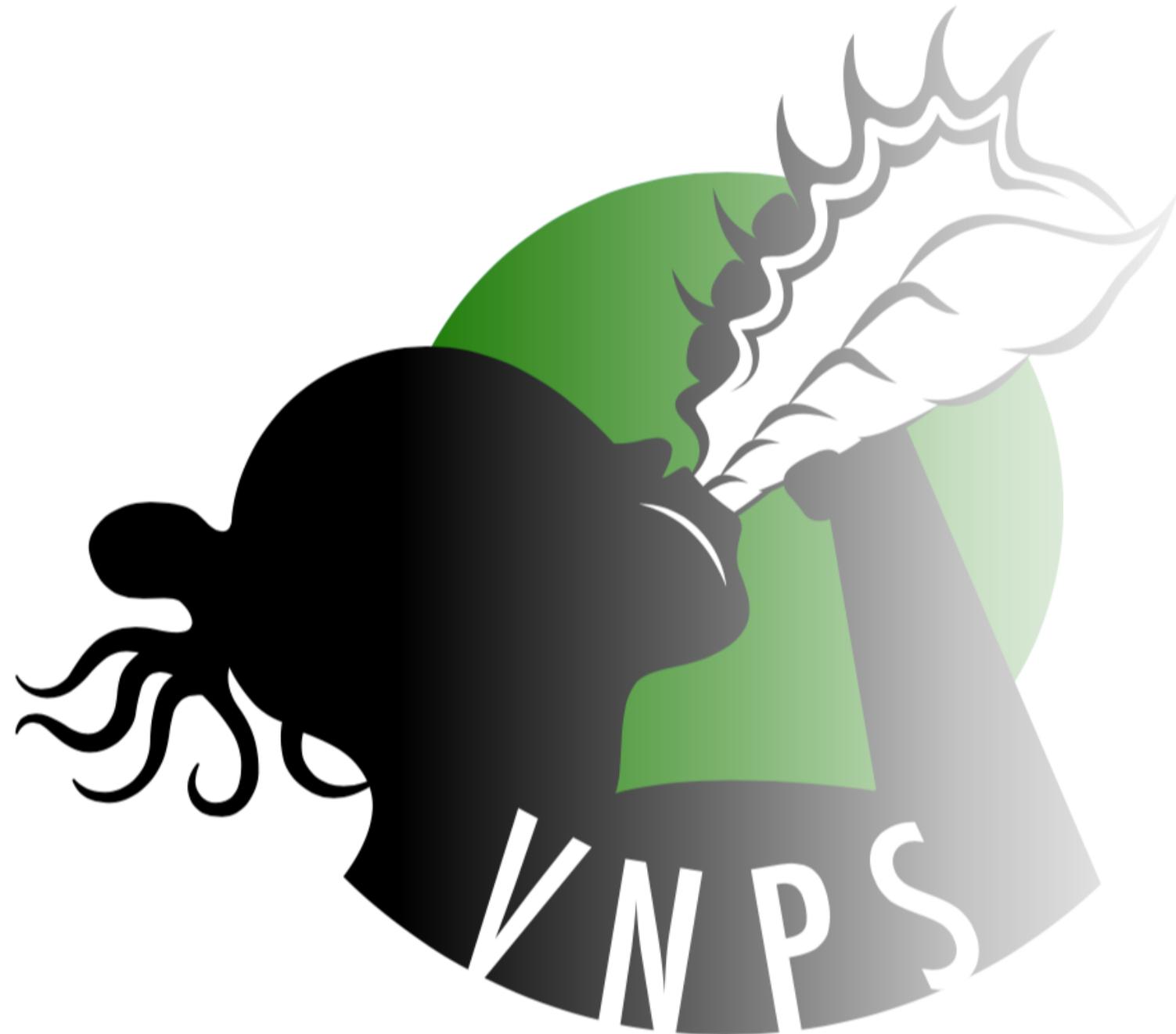


- *L'operation security* ou sécurité opérationnelle (OpSec) est le fait de protéger son identité lors d'une investigation.
- Au préalable, il convient de déterminer le **modèle de menace** : contre qui ou quoi souhaite-t-on se prémunir et pourquoi.
- Cela permettra d'ajuster les mesures de protection à mettre en œuvre : utilisation d'un compte factice (*sock puppet*) ou d'un VPN, par exemple.

- **Les incontournables de l'hacktiviste :**

Mieux se protéger (sécurité numérique) & mieux s'informer (investigation en ligne) :

<https://osintfr.com/fr/accueil/> | <https://nothing2hide.org/fr>



CONCLUSION

- Suite à nos démissions respectives de la logistique humanitaire (ONU) et de la banque d'investissement (HSBC), ma compagne et moi-même avons fondé l'association « **Vous N'êtes Pas Seuls** » :
 - <https://vous-netes-pas-seuls.org/>
- Elle a pour but d'accompagner des salariés souffrant d'une fracture entre leur travail et leurs valeurs, d'accumuler des connaissances d'initiés sur les nuisances de leurs secteurs, de diffuser les témoignages de leur rupture, tout en s'inspirant des alternatives prometteuses existantes.

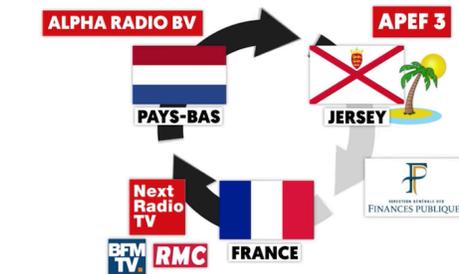
CONCLUSION

- Le média activiste « **Lanceur d’alerte** » fondé par Maxime Renahy — ancien administrateur de fonds dans un cabinet d’avocats d’affaires à Jersey — et son frère Alexandre — journaliste citoyen et militant écologiste — sont nos partenaire privilégiés pour accompagner de potentiels futurs lanceurs d’alerte.
 - <https://lanceuralerte.org/>



CONCLUSION

- En combinant « OSINT » et fuite d'informations sensibles, ils ont déjà provoqué plusieurs actions en justice pour forcer multinationales (Lactalis, Charal, etc.) et milliardaires (B. Arnault) à publier leurs comptes afin de lutter contre leurs évasions fiscales.



Juillet 2019

L214 fait plier le groupe Bigard-Socopa-Charal

Une impunité qui prend fin

Septembre 2019

Pourquoi la Confédération paysanne attaque Lactalis en justice

Le syndicat agricole demande au géant de l'agroalimentaire de publier ses comptes annuels d'ici un mois.

Octobre 2019

Le Club de Mediapart

Enquête contre les grandes fortunes : jusqu'où ira Le Média ?

... Maxime et Alexandre Renahy, annonce une spectaculaire opération ... Lagardère, Bernard Arnault ou encore Marc Ladreit de Lacharrière.

Octobre 2019

AVEZ-VOUS DES QUESTIONS ?

