*Séminaire Aristote*

Données de la recherche : Nature, Cultures & Valeurs

*Sorbonne Université, 18 oct 2023*

# COÛT DES DONNÉES (ENVIRONNEMENT ET ÉCONOMIQUE)

## Regard d'une chercheuse

Cécile CALLOU, MC MNHN

UAR BBEES & UMR AASPE (CNRS-MNHN)

# En guise d'introduction

Enseignant-chercheur en Archéozoologie au sein de l'unité « Archéozoologie, archéobotanique : sociétés, pratiques et environnements (AASPE) » (CNRS-MNHN)

- Production, Traitement et analyse des données
- Recherche de financements

Responsable scientifique de collection et administratrice de la base de données « Zoologie Mammifères et Oiseaux – Anatomie comparée »

- Gestion
- Conservation et diffusion des données

→ Données : haute valeur scientifique

# En guise d'introduction

Directrice de l'unité « Bases de données sur la Biodiversité, Écologie, Environnement et Sociétés (BBEES) » (CNRS-MNHN)

- Accompagnement des acteurs de la recherche (T, I, C & EC) sur l'ensemble du cycle de vie de la donnée

Recommandations sur la constitution des corpus et le traitement des données, le choix des outils (Open Source), la structuration des données, les métadonnées, la mise en place de plan de gestion de données, les chartes d'utilisation des données, l'adoption de licences libres, etc. … mais aussi le choix d'hébergement ou de stockage.

Une attention particulière est portée aux standards et normes en vigueur, au choix de référentiels et thesaurus partagés, aux réglementations concernant la propriété intellectuelle, RGPD et Intégrité scientifique.

# En guise d'introduction

Directrice de l'unité « Bases de données sur la Biodiversité, Écologie, Environnement et Sociétés (BBEES) » (CNRS-MNHN)

- Mutualisation d'applications et création d'outils communs pour la publications et la diffusion des données

Partenariat avec le CC-IN2P3 (UAR6402 CNRS)

# Tenir compte du coût de gestion de la donnée

Pour l'estimer, un outil : le plan de gestion des données (PGD)
Évaluer les coûts en amont permet de les financer, mais aussi de les réduire à terme en adoptant d'emblée des choix et des pratiques de gestion de données conformes aux principes FAIR.

https://doranum.fr/enjeux-benefices/le-cout-de-la-gestion-des-donnees_10_13143_hch2-h207/

# Tenir compte du coût de gestion de la donnée

| Modèle d'estimation des coûts de gestion des données | | | | |
|---|---|---|---|---|
| *Titre du projet :* | | | | |
| **Phases du cycle de vie** | **Activité** | **Quantité** | **PU HT** | **Total** |
| ***Données collectées - produites*** | Acquisition de jeux de données externes | | | 0,00 |
| | Mise en forme et organisation * | | | 0,00 |
| | Logiciel pour traitement des données | | | 0,00 |
| | Formation(s) | | | 0,00 |
| | Transcription | | | 0,00 |
| | Traduction | | | 0,00 |
| ***Métadonnées - Documentation*** | Frais de personnel pour la description des données * | | | 0,00 |
| | Frais de personnel pour la documentation * (contexte, méthodologie) | | | 0,00 |
| ***Stockage des données - sauvegarde*** | Espace de stockage supplémentaire | | | 0,00 |
| | Logiciels ou matériel supplémentaires | | | 0,00 |
| | Achat d'un sytème de cryptage | | | 0,00 |
| | Coût de l'accès sécurisé pour des personnes externes | | | 0,00 |
| ***Contrôle - qualité - responsabilité*** | Coût Responsable du PGD * | | | 0,00 |
| | Coût Responsable de la qualité des données * | | | 0,00 |
| | Conseil juridique (droit d'auteur) | | | 0,00 |
| ***Archivage - partage*** | Frais de personnel pour la conversion dans un format standard ou ouvert | | | 0,00 |
| | Logiciels ou du matériel supplémentaires pour la conversion | | | 0,00 |
| | Coût du dépôt dans l'entrepôt | | | 0,00 |
| | Nettoyage des données * | | | 0,00 |
| | Numérisation | | | 0,00 |
| ***Cas particulier : anonymisation*** | Frais de personnel pour anonymiser les données * | | | 0,00 |
| | Logiciels | | | 0,00 |
| ***Cas particulier : préservation à long terme*** | Nbre ou poids des fichiers à conserver au-delà de 30 ans | | | 0,00 |
| | | | **Total général** | 0,00 |

## Prise en charge par financeur (exemples):

- **Frais de personnel**
- **Coûts des instruments, du matériel et des consommables scientifiques** : coût d'acquisition ou de location des instruments ou des matériels et des consommables scientifiques
- **Coûts des bâtiments et des terrains**
- **Licences, cession de brevet, marque, logiciel, base de données, droit d'auteur etc.**
- **Coûts des prestations de services**
- **Frais généraux** : frais de mission, déplacement des personnels permanents ou temporaires affectés au projet, frais de réception et organisation de séminaires/colloques en lien avec le projet
- **Frais de publication** ou APC (Article Processing Charges)

https://doranum.fr/enjeux-benefices/le-cout-de-la-gestion-des-donnees_10_13143_hch2-h207/

# Coût *versus* Valeur

→ Coût extrêmement variable selon les communautés

- Elément de réflexion : Les données accessibles dans le domaine environnemental. Quels freins? (AllEnvi, *GT Infrastructures, systèmes d'observation et services d'information environnementale et climatique*, Paris, https://www.allenvi.fr/seminaire-gerer-les-donnees-environnementales-mai-2014/)

  Quels moyens (financiers et humains) ont été nécessaires pour mettre en place ces outils ; Description des modalités d'accès (Partage des données) ; quels moyens déployés pour assurer la maintenance et l'accès; Comportement de la communauté scientifique concernée

→ Activité de recherche s'inscrivant dans un cadre national et européen ayant transformé les pratiques de diffusions et de publications des données

→ Valeur stratégique des données --- Gouvernance des données

# CHANGEMENT RÉGLEMENTAIRE POUR LA SCIENCE

**1978**

LOI n°78-753 du 17 juillet 1978
portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal
(loi CADA)

**2015**

LOI n°2015-1779 du 28 décembre 2015
relative à la gratuité et aux modalités de la réutilisation des informations du secteur public
(loi Valter)

**2016**

REGLEMENT (UE) n°2016-679 du 27 avril 2016
général sur la protection des données

LOI n°2016-1087 du 7 août 2016
pour la reconquête de la biodiversité, de la nature et des paysages

LOI n°2016-1321 du 7 octobre 2016
pour une République numérique
(Loi Lemaire)

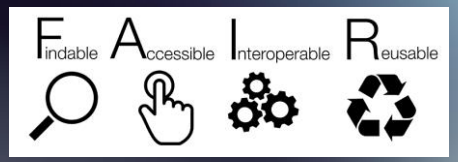
**2018**

**2020**

**2021**

DECRET n°2021-1572
du 3 décembre 2021
relatif au respect des **exigences de l'intégrité scientifique** par les établissements publics[…]ayant pour **activité principale la recherche**

> *"La France s'engage pour que les résultats de la recherche scientifique soient ouverts à tous, chercheurs, entreprises et citoyens, sans entrave, sans délai, sans paiement."*



## Loi pour une République numérique dite « Loi Lemaire » (n° 2016-1321 du 7 octobre 2016)

- **Ouverture par défaut des données des administrations publiques**. L'objectif fixé par la loi est une publication des principaux documents administratifs, puis de **tous les documents qui revêtent un intérêt économique**, social, sanitaire ou environnemental. *[dont les données de la recherche]*
  - Décret 2016-1922 du 28 dec 2016 relatif à la publication en ligne des documents administratifs
  - Décret 2017-638 du 27 avril 2017 relatif **aux licences de réutilisation à titre gratuit des informations publiques**
- La gratuité des échanges de données entre administrations de l'Etat **(janvier 2017)**
- **Service public de la donnée (art. 14)**. Il vise à mettre à disposition, en vue de faciliter leur réutilisation, des jeux de données de référence, c'est-à-dire **ceux qui présentent le plus fort impact économique** et social, pour que les entreprises puissent les réutiliser pour leurs services
  - Décret 2017- 331 du 14 mars 2017 relatif au service public de mise à disposition des données de référence
- **Le libre accès aux publications scientifiques de la recherche publique (art. 30)**, grâce au droit dévolu aux chercheurs de diffuser leurs articles après une courte période d'embargo de 6 à 12 mois, et ce quel que soit le contrat entre le chercheur et l'éditeur de la revue publiant l'article

# DES POLITIQUES D'OPEN ACCESS, D'OPEN DATA ET DE SCIENCE OUVERTE…

Les politiques publiques de science ouverte prennent **racine au niveau international** au début des années 90. Elles sont ensuite **portées par des initiatives scientifiques intercommunautaires,** sont **institutionnalisées par l'Europe** qui se positionne en premier plan et sont aujourd'hui **déclinées au niveau national et ministériel** :

**Open Access** → **Open Data** → **Science Ouverte** → **Gestion de la donnée, des algorithmes et des codes sources**

| | *1990* | *2000* | *2005* | *2010* | *2015* | *2020* |
|---|---|---|---|---|---|---|
| International | **1993** ONU - Création de la convention sur la diversité biologique<br>**1993** Déclaration de l'UNESCO et du CIUS sur la science et l'utilisation du savoir scientifique | **2002** Initiative Budapest pour l'accès libre<br>**2003** Déclaration de Bethesda pour l'édition en libre accès<br>**2003** Déclaration de Berlin sur le libre accès en sciences exactes, sciences de la vie, humaines et sociales<br>**2004** OCDE - Premiers travaux sur la Science ouverte | **2006** OCDE - Recommandation concernant l'accès aux données de recherche financées par des fonds publics<br>**2009** UNESCO - Charte sur la conservation du patrimoine numérique | **2010** ONU/CDB - Protocole de Nagoya, Plan stratégique pour la diversité biologique 2011-2020 et les Objectifs d'Aichi<br>**2012** UNESCO- Principes directeurs pour le développement et la promotion du libre accès. | **2015** Les 17 objectifs de développement durable de l'ONU<br>**2017** UNESCO - Les recommandations sur la Science et la recherche scientifique<br>**2019** UNESCO - Les recommandations sur les ressources éducatives libres (REL) | **2020** ONU - Un Cadre mondial de la biodiversité pour l'après 2020<br>**2021** OCDE - Révision des recommandations sur l'accès aux données de recherche<br>UNESCO - Recommandations sur la Science ouverte |
| Européen | **2000** Création de l'Espace européen de la recherche | **2006** European Research Council (ERC), déclaration du Conseil scientifique pour l'accès libre | **2008** Open Access Pilot | **2010** Stratégie numérique de l'Europe : Programme Horizon 2020<br>**2012** CE - Recommandation relative à l'accès aux informations scientifiques et à leur conservation<br>**2013** Horizon 2020, le programme de financement de la recherche et de l'innovation<br>**2015** CE - La stratégie pour un marché unique numérique | **2016** CE- Recommandations aux différentes parties prenantes de l'Open Science et la transition vers un système global de science ouverte<br>**2018** CE - Procédure de revitalisation de l'Espace Européen de la Recherche et révision de la recommandation "Accès et préservation de l'information scientifique",<br>**2018** Déclaration de Carlos Moedas, "Le nuage européen de la science ouverte (EOSC), | **2020** CE - Nouvelle stratégie de biodiversité pour 2030<br>**2020** CE - Horizon Europe, le programme de financement de la recherche et de l'innovation<br>**2022** CE - Data Governance Act - Digital Service Act – Data Act – Cyber resilience Act |
| National | | | | **2013** Mission Etalab<br>**2013** Lancement des politiques du libre accès à l'horizon 2020 | **2016** MESR – Loi pour une République numérique<br>**2016** MTE - Loi pour la reconquête de la biodiversité, de la nature et des paysages<br>**2018** MESR- Stratégie nationale pour l'intelligence artificielle<br>**2018** MESR- PNSO1 | **2019** Rattachement d'Etalab au sein DINUM<br>**2020** Rapport Bothorel sur la politique de la donnée, des algorithmes et codes sources + circulaire<br>**2020** MESR - Deuxième PNSO<br>**2021** MTE Feuille de route de la donnée, des algorithmes et des codes sources<br>MESR – Feuille de route des données des algorithmes et des codes sources |

Les politiques publiques de science ouverte reposent sur un triple registre de valeurs et de réflexion :

- Une question **d'ouverture des données scientifiques**
- Guidée par une prise de conscience environnementale
- Portée par une ambition d'innovation et de gain sociétal

## Règlement sur la gouvernance des données (« Data Governance Act  »), **24 septembre 2023**

- Vise à favoriser et encadrer le partage de données (personnelles et non personnelles) et crée notamment un statut réglementé **d'intermédiaire de données** fondé sur la confiance (**tiers de confiance**), organisant la mise à disposition des données de façon neutre afin d'éviter des conflits d'intérêts.
- Introduit aussi le concept **d'altruisme des données** : « les individus et les entreprises qui donnent leur consentement ou la permission de mettre à disposition des données qu'elles génèrent — volontairement et sans récompense — pour être utilisées dans l'intérêt public ». Ce statut répondant à un ensemble de critères stricts cherche à motiver les acteurs privés pour qu'ils partagent leurs données.
- Prévoit la mise en place du **règlement sur les données** (« **Data Act** »), entrée en vigueur en 2025, dont les objectifs sont :
    - faciliter le partage des données émanant d'objets connectés en contrepartie d'une compensation juste, équitable et proportionnée ;
    - autoriser les organismes publics des États membres et les institutions de l'Union européenne à avoir un accès ainsi qu'un droit d'utilisation de ces données à condition de pouvoir justifier d'un besoin exceptionnel ;
    - établir des règles de coopération internationale en matière de transmission des données non personnelles ;
    - lutter contre les accès illicites à ces données et notamment contre les intrusions de gouvernements de pays tiers.

    → marché numérique unique des données.

**À consulter :** « Mise en œuvre du règlement sur la gouvernance des données (Data Governance Act) : les enjeux économiques » (https://www.cnil.fr/) ;  https://www.vie-publique.fr/en-bref/290144-marche-unique-des-donnees-vers-ladoption-du-data-act

Commission européenne

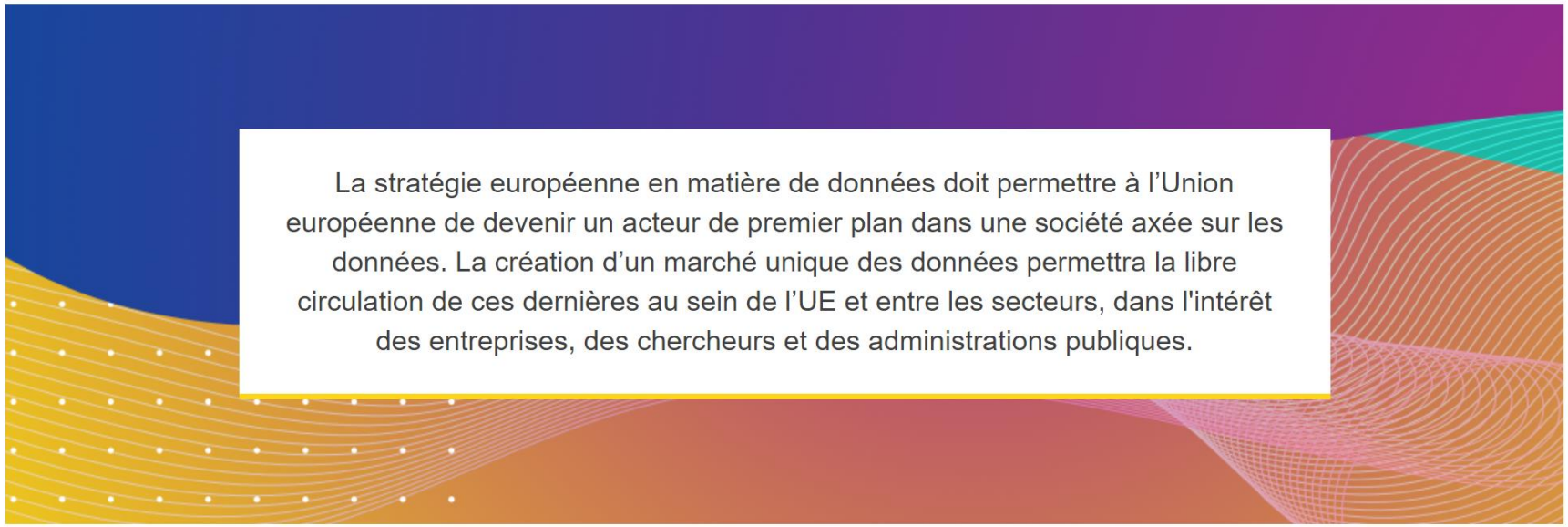
français

Rechercher

Accueil > Aspects stratégiques et politiques > Priorités > Une Europe adaptée à l'ère du numérique > Stratégie européenne pour les données

# Stratégie européenne pour les données

Faire en sorte que l'UE devienne un modèle et un acteur majeur d'une société dont les moyens d'action sont renforcés par les données.




https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr

# Projections pour 2025

530 %

**d'augmentation du volume mondial de données**

de 33 zettaoctets en 2018 à 175 zettaoctets

829 milliards d'euros

**Valeur de l'économie fondée sur les données dans l'EU-27**

contre 301 milliards d'euros (2,4 % du PIB de l'UE) en 2018

10,9 millions

**de professionnels des données dans l'EU-27**

contre 5,7 millions en 2018

65 %

**Pourcentage de la population de l'UE possédant des compétences numériques de base**

contre 57 % en 2018

https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr

Zo=$10^{21}$

## Que faire ?

S’inscrire dans une démarche plus large portée par les institutions (Stratégie de gouvernances des données: technique, organisationnel et politique),

en respectant les usages de la communauté disciplinaire (collecte, stockage, valorisation) tout en les faisant évoluer (décloisonnement nécessaire).

- Présence de freins
  - Prise de conscience insuffisante sur l’importance de la donnée
  - Manque de culture de la donnée
  - Manque de compétences (métiers de l’informatique)
  - Multiplicité des outils
  - Non conscience du coût/valeur/impact de la donnée : économique et environnemental

# LES ENJEUX ENVIRONNEMENTAUX ISSUS DE L'EXPLOITATION DES DONNÉES

Si les données permettent d'accroître la connaissance sur l'environnement et de mesurer l'impact écologique, chaque traitement sollicite des éléments fonctionnels (du smartphone au serveur, en passant par les réseaux des opérateurs) qui ont un impact environnemental fort.

En 2019, le numérique mondial représentait un 7ème continent de la taille de 2 à 3 fois celle de la France.

Contribution :

- Consommation d'énergie primaire (EP) : 4,2 %
- Emissions de gaz à effet de serre (GES) : 3,8 %
- Consommation d'eau (eau) : 0,2 %
- Consommation d'électricité (Elec.) : 5,5 %

Rapporté à des usages quotidiens, cela revient à :

- GES : 1,5 milliard de salariés français allant travailler pendant 1 an
- Eau : 242 milliards de packs d'eau minérale (9 litres)
- Electricité : 82 millions de radiateurs électriques (1000 Watts) allumés en permanence.

DEEE=Déchets équipements électriques et électroniques

L'empreinte environnementale du numérique

# Etude ADEME – Arcep sur l'empreinte environnementale du numérique en 2020, 2030 et 2050

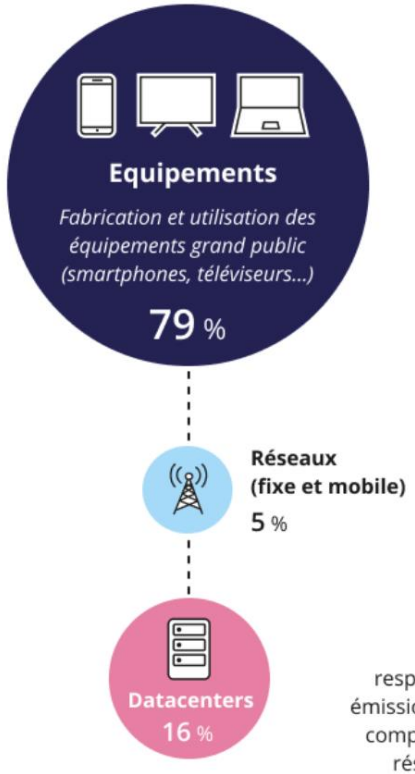
Mis à jour le 13 mars 2023

https://www.arcep.fr/la-regulation/grands-dossiers-thematiques-transverses/lempreinte-environnementale-du-numerique/etude-ademe-arcep-empreinte-environnemental-numerique-2020-2030-2050.html

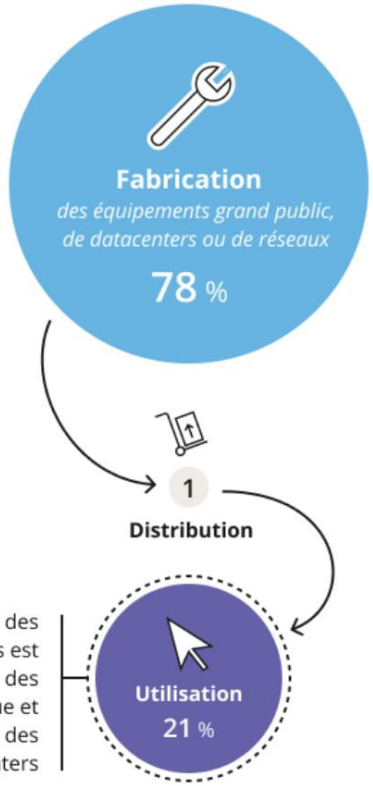
# EMPREINTE CARBONE

## L'empreinte carbone du numérique dépend essentiellement des équipements et de leur fabrication

Répartition de l'empreinte carbone du numérique en 2020 par composantes du numérique (%)




Répartition de l'empreinte carbone du numérique en 2020 par phase du cycle de vie (%)



L'empreinte environnementale du numérique

# Etude ADEME – Arcep sur l'empreinte environnementale du numérique en 2020, 2030 et 2050

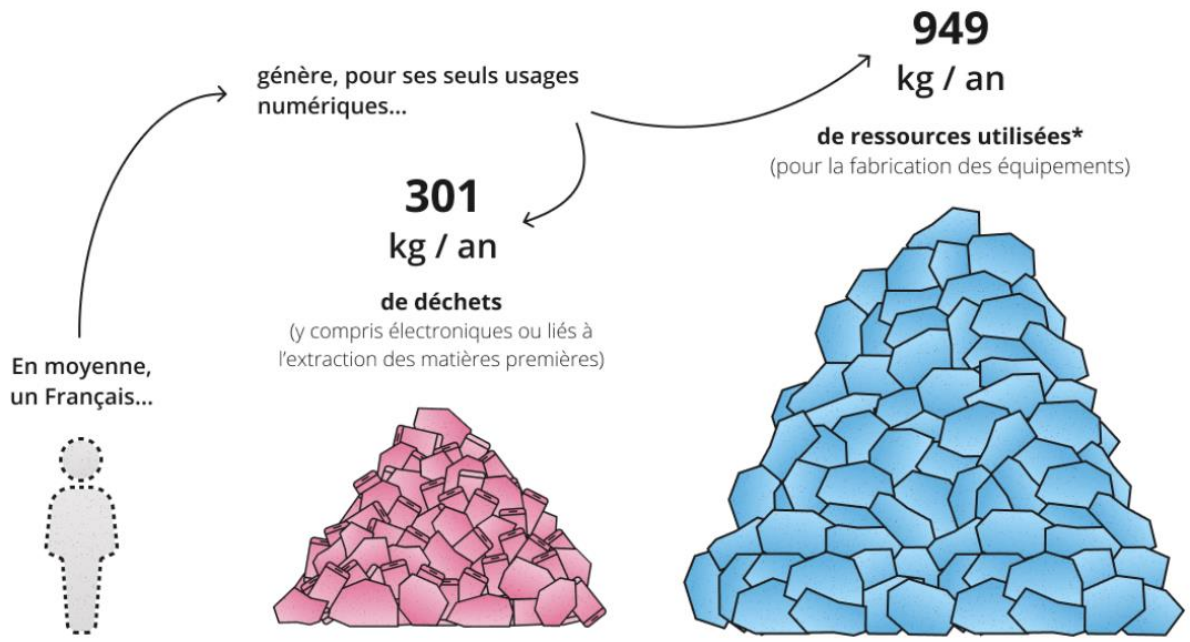
Mis à jour le 13 mars 2023

https://www.arcep.fr/la-regulation/grands-dossiers-thematiques-transverses/lempreinte-environnementale-du-numerique/etude-ademe-arcep-empreinte-environnemental-numerique-2020-2030-2050.html

# CONSOMMATION DES RESSOURCES

## Près d'une tonne de matériaux utilisés par personne chaque année pour nos usages numériques

Quantité de ressources utilisées ou de déchets produits chaque année pour répondre aux usages numériques d'une personne vivant en France en 2020

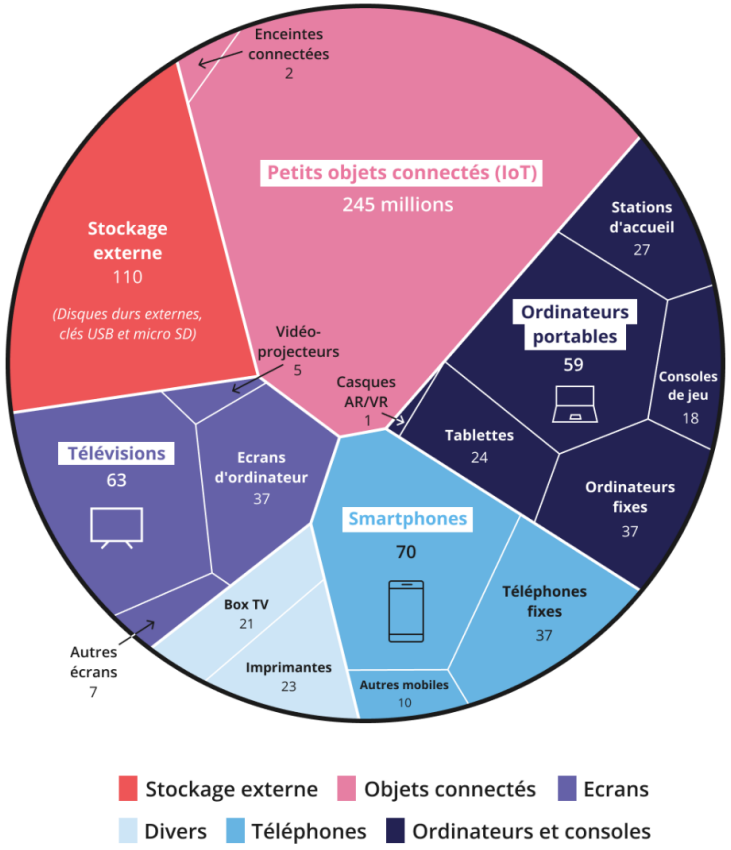



* comprenant ressources abiotiques (matériaux, énergie fossile...), biomasse, déplacements de terre et l'eau.

# IMPACT SELON LE TYPE D'EQUIPEMENT

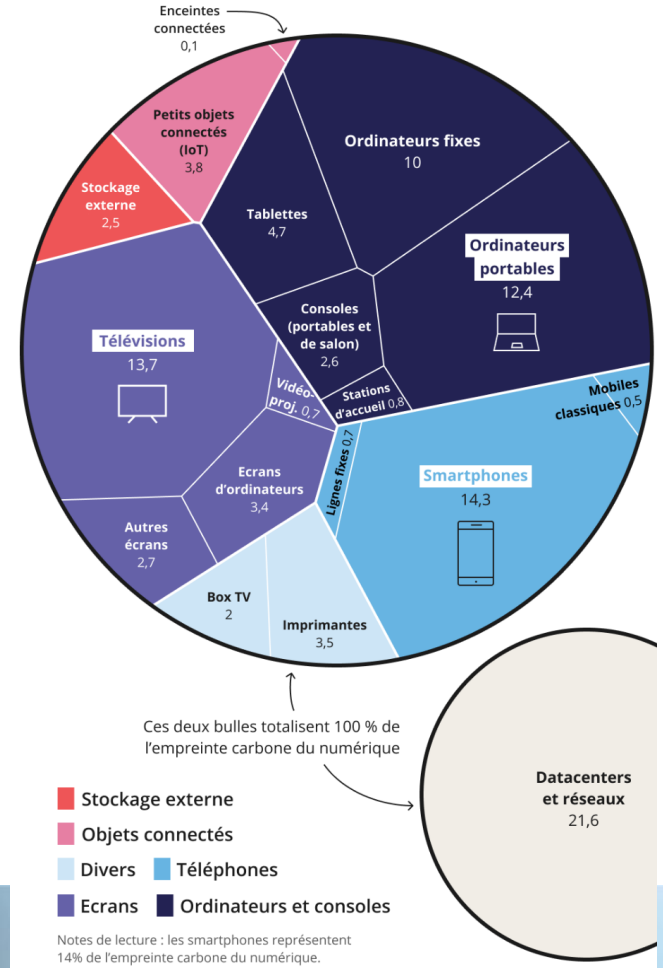
## Plus de 800 millions* d'équipements utilisateurs en 2020 en France

Estimation du nombre de terminaux utilisateurs utilisés en France pour des usages personnels et professionnels (en millions)




## Les smartphones et les téléviseurs, les deux principaux moteurs des émissions du numérique

Part de chaque type d'équipement (sur tout le cycle de vie) dans l'empreinte carbone du numérique en 2020, comparée à celle des centres de données et réseaux (%)




Notes de lecture : les smartphones représentent 14% de l'empreinte carbone du numérique.

## Impact environnemental- vers une prise de conscience

GDR (CNRS, INRAE, ADEME, INRIA, SorbonneU) Labos 1point5

GDS (CNRS-INS2I et INEE)) EcoInfo Groupement de service

POUR UNE INFORMATIQUE ÉCO-RESPONSABLE

ANF 2023 Impacts environnementaux du numérique : les 8R (20-24 nov 2023) – Ressentir, Réfléchir, Réduire, Refuser, Réparer, Ré-utiliser, Recycler, Ralentir)

Création d'indicateurs du numérique dans les institutions

# MERCI !

**C. CALLOU**
**SÉMINAIRE ARISTOTE**
**SORBONNE-UNIVERSITÉ, 18 OCT 2023**