



MINISTÈRES
TRANSITION ÉCOLOGIQUE
COHÉSION DES TERRITOIRES
TRANSITION ÉNERGÉTIQUE
MER

Liberté
Égalité
Fraternité



Feuille de route ministérielle de la donnée,
des algorithmes et des codes sources

Facilitateur de vos transitions
#CMVRH



" Les données, pourquoi en parle-t-on ? "
1^{er} webinaire d'un cycle sur la culture de la donnée
Jeudi 26 janvier 2023 de 9h30 à 12h15

Partie 1. Pourquoi parle-t-on des « données » depuis quelques années au sein du pôle ministériel ?

Intervenante : Hélène BEGON - CGDD/SRI/Ecolab

Auteure de : La transformation numérique des administrations, La documentation française, août 2021

Support préparé avec les collègues de l'équipe data-IA de l'Ecolab du Commissariat général au développement durable



On attend de nous, agents publics :

- De continuer à créer de la donnée de qualité
- De continuer à nous appuyer sur de la donnée de qualité pour mettre en œuvre un service public de qualité
- De continuer à garantir un droit d'accès aux documents administratifs
- De continuer à protéger les secrets (médical, des affaires, de la sécurité nationale...)
- De continuer à protéger le patrimoine de savoir lié à la donnée (Archives)

Mais on attend aussi de nous, agents publics, **dans une société désormais transformée par les usages numériques**

- De partager et faire partager beaucoup plus cette donnée de qualité pour qu'elle serve à l'économie (croissance, compétitivité), aux citoyens (transparence, débat démocratique), à la recherche, ...
- D'utiliser beaucoup plus les outils qui enrichissent la donnée, afin d'offrir un meilleur service public
 - Datavisualisations plus instructives
 - Croisements de données faisant mieux apparaître les phénomènes
 - Plateformes internet de services liés à la donnée, outils permettant aux administrations de s'échanger les données
 - Usages de l'intelligence artificielle, etc.
- De protéger les citoyens, l'économie, le contrat social, contre des nouvelles formes d'usages abusifs des données
 - Protection des données personnelles, des données commerciales, en ligne
 - Limitation des conditions d'utilisation des outils d'intelligence artificielle
 - Régulation des usages d'internet, de la puissance de la « Big tech », etc.

La « politique publique de la donnée » est en effet indissociable des effets grandissants de l'ère numérique

Une prodigieuse progression du numérique :

- Technique : internet, capacités de stockage, puissance de calcul, multiplication des smartphones et autres objets connectés, amélioration des réseaux d'échange (technologie 5G, déploiement de la fibre), temps réel, ...
- Economique et sociétale : poids des entreprises de la « big tech », taux d'utilisation d'internet par la population (92% en Europe, 63% dans le monde)

La création d'une masse et d'une variété toujours plus considérables de données, sur lesquelles de grands opérateurs numériques assoient leur richesse et leur potentiel (« *La donnée est le pétrole du 21^{ème} siècle* »)

- ➔ *90 % des données récoltées depuis le début de l'humanité auraient été générées durant les 2 dernières années (IBM en 2019)*
- ➔ *La valeur de l'économie fondée sur les données dans l'EU-27 serait de 829 milliards d'euros en 2025 (Commission européenne)*
- ➔ *Le volume mondial des données devrait augmenter de 530 % d'ici à 2025 (Commission européenne)*

Les techniques de gestion de la donnée progressent pour gérer (et accroître !) cette « big data » (donnée massive) :

- outils de visualisation, de modélisation (jumeaux numériques)
- science algorithmique, science des données
- services de data science accessibles à distance (*cloud*)
- performance des moteurs de recherche
- reconnaissance automatisée des images ou du langage
- autres multiples usages de l'intelligence artificielle, ...

*De facto, parler aujourd'hui des données, c'est parler du numérique
Même si les deux notions ne sont pas équivalentes, elles s'enrichissent l'une l'autre de façon très étroite*

Peut-on définir « une donnée » ?

CONSTRUITE, CONTINGENTE, UNE DONNÉE N'EST PAS « LA » VÉRITÉ
MAIS ELLE EST CRÉÉE POUR COMPRENDRE LE RÉEL

Une donnée est :

- une **brique, une base de la connaissance** : c'est une observation, une mesure, une description, un résultat, un enregistrement, une géolocalisation, un identifiant (*n° de sécurité sociale*)
- **construite, signifiante** : elle existe dans un certain contexte, elle obéit à un référentiel (d'origine humaine), elle doit pouvoir être comprise et interprétée par ceux qui connaissent son référentiel
- et **représentée sous une forme conventionnelle** destinée à faciliter son traitement : en kilomètres, en pixels, en coordonnées géographiques, ...

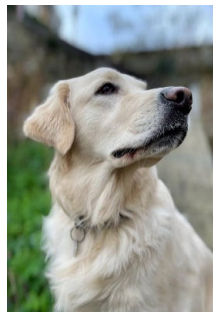
Une donnée permet de concevoir, représenter, comprendre des phénomènes réels ou virtuels et, avec les multiples progrès de son traitement, d'organiser de plus en plus finement la vie humaine.

Les données sont l'objet roi de la démarche scientifique.

Ce sont les scientifiques qui sont à l'origine de l'ouverture et de l'échange des données, mouvement dit d' « open data » (*données ouvertes*), bien avant l'ère numérique.

La démarche scientifique élabore des outils pour créer et utiliser les données, mais aussi et tout autant pour « s'en méfier » et ne pas leur « faire dire » ce qu'elles ne « disent » pas.

Peut-on définir « une donnée » ?



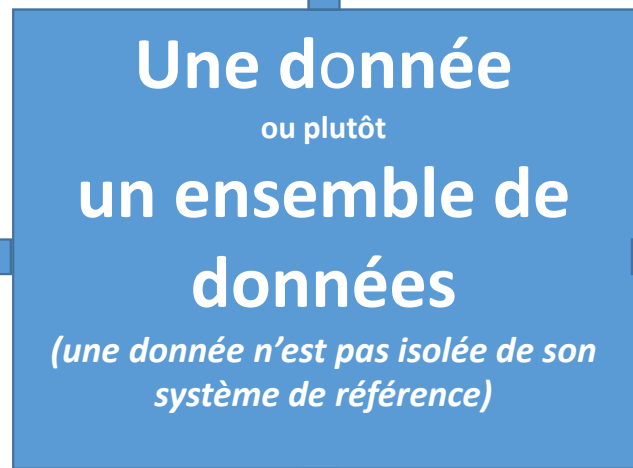
Une photo

Accès aux documents : [https://www.km\[...\]kmae/2022023/pdf](https://www.km[...]kmae/2022023/pdf)

Un document texte



Les informations fournies en temps réel par un capteur



Un post sur un réseau social



Une donnée personnelle



Rapport d'évaluation
environnementale
du programme régional
forêt bois de Nouvelle-Aquitaine
Résumé non technique



Centre d'étude et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement (Cerema)

Un rapport



Un fichier vidéo

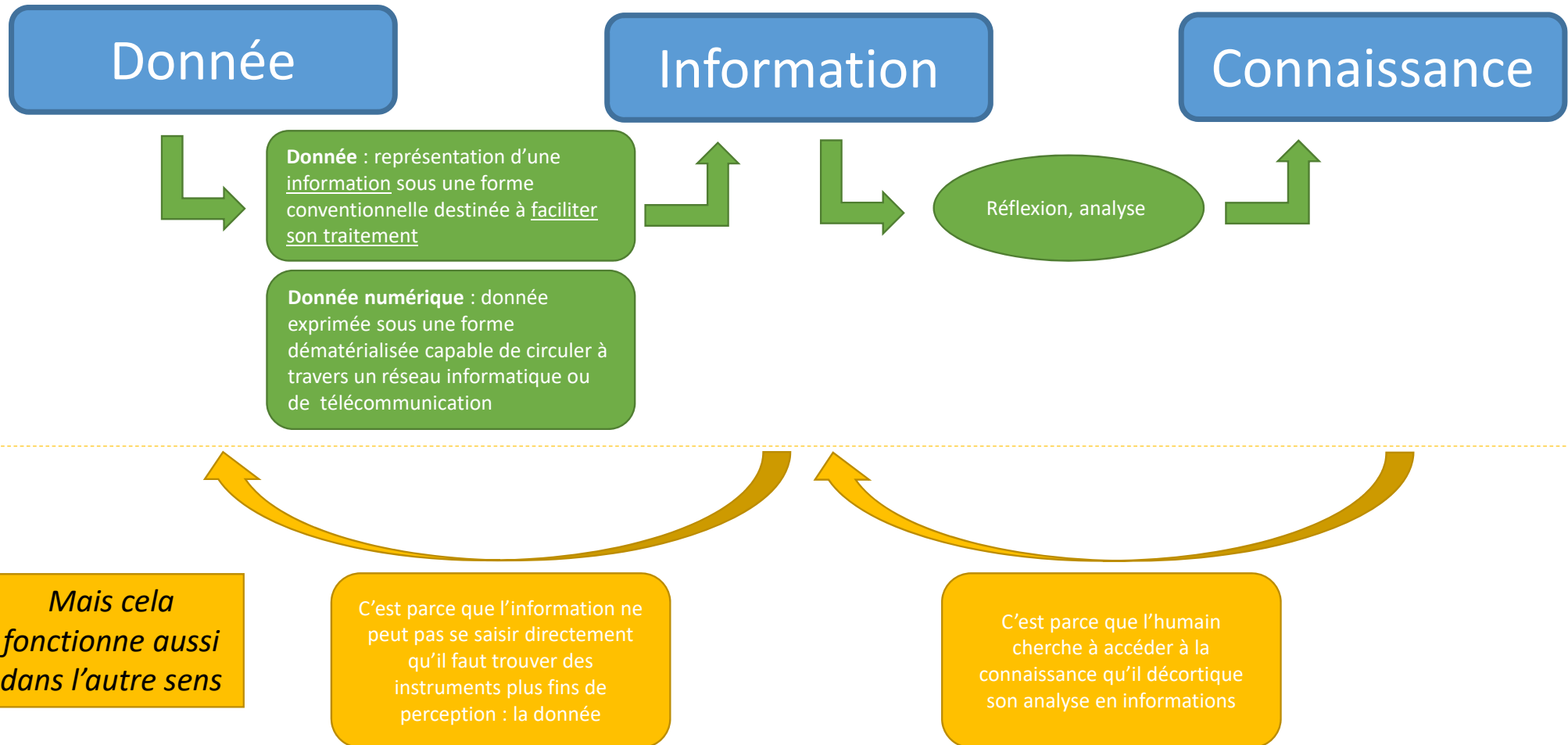
	Total	Sexe		Âge		
		Femmes	Hommes	15 à 24 ans	15 à 64 ans	65 ans et plus
Actifs	29 556	14 224	15 332	2 766	26 441	349
Employés	26 584	12 822	13 761	2 085	24 158	341
Chômeurs	2 972	1 402	1 571	681	2 282	9
Inactifs	23 286	13 334	9 952	4 723	6 960	11 603
Total	52 842	27 558	25 284	7 489	33 401	11 952

La population active française en 2016 (en milliers, personnes de 15 ans ou plus)

Des chiffres dans un tableau Excel

Peut-on définir « une donnée » ?

On a coutume de faire le lien suivant :



Peut-on définir « une donnée » ?

La donnée n'est pas et ne peut pas être une affaire réservée aux « techniciens »
Il est utile d'acquérir une culture de la donnée

Pour organiser un bon usage des données, il faut une collaboration étroite entre les experts de trois champs :

- la donnée
- l'informatique, le web
- et évidemment les « métiers »

La donnée est une notion imprécise et multicartes
Il ne faut pas s'arrêter au terme de « donnée » mais comprendre la logique qui la sous-tend dans notre contexte professionnel : mieux comprendre le réel et travailler avec les standards techniques modernes

Il y a la donnée pour le juriste, pour l'informaticien, pour le scientifique de la donnée, pour le statisticien, pour le géomaticien, pour l'économiste...*
Ils ne s'intéresseront pas aux mêmes enjeux et n'utiliseront pas les mêmes outils

* NB : existe aussi en féminin 😊

Dans les textes on va trouver les termes « donnée », « base de données », « information publique », « document administratif », « traitement algorithmique »...

Il ne faut pas s'en tenir au pied de la lettre car « la donnée » est un peu dans tout

Il y a « donnée » (peu utile) et « donnée » (souhaitable)

Enrichissements d'une donnée :

- Donnée brute, utile uniquement à son créateur (et encore, pas longtemps...) surtout si elle est non numérique
(Encore que les chercheurs peuvent la préférer car elle n'est pas polluée d'interprétations, ou contrainte par un modèle conceptuel)
- ↗ Assortie de **métadonnées** (date, lieu, auteur, géoréférencement, licence de réutilisation, etc.)
- ↗ De **qualité** : récente, comparable dans le temps et homogène dans l'espace, croisée avec des « données de référence » (ex : base adresses nationale, code de la commune) etc etc. ; *il y a de très nombreux critères de qualité de la donnée*
- ↗ Inscrite dans un **schéma** partagé de données qui précise les formats des données, les métadonnées à associer, ...
- ↗ **Ouverte** en ligne, facile à trouver (inscrite dans des catalogues) et à comprendre, purgée des risques juridiques (donnée personnelle, droits de propriété, ...) - « *Toute mise à disposition (...) sous forme électronique (...) se fait dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé* » (L.300-4 CRPA)
 - ↗ Ou au contraire **protégée**, cryptée...
- ↗ **Partagée** au sein d'un « Commun numérique » (communauté d'acteurs qui l'enrichissent collectivement)
- ↗ **Projetée** sous forme d'animation vidéo, de data-visualisation, de podcast, de publication...
- ↗ **Organisée** dans une base de données, un système d'information, une plateforme internet gérés en lien avec les besoins des utilisateurs
- ↗ Assortie là d'**outils** pour la traiter, enrichir l'information qu'elle contient potentiellement : moteur de recherche, croisements de données, visualisations (schémas), visualisations multi-couches (cartes), algorithmes aidant à la prise de décision, puissance de calcul associée à l'outil, ...
- ↗ Largement et **massivement** accessible : API, croisements automatiques (type « Dites-le nous une fois »), haut débit de téléchargement...
- ↗ Pouvant servir à l'entraînement d'outils d'**intelligence artificielle**, et encore mieux si de la donnée est partagée entre de nombreux acteurs (espaces communs thématiques de données : ex *Health Data Hub*)
- ↗ Gérée et utilisée de façon à minimiser ses effets **environnementaux**
- ↗ Pensée dans un **cycle de vie** complet, y compris sa destruction le cas échéant.

Mais ceci se brouille avec les outils qui extraient de la donnée à partir de sources dispersées, ou permettent de stocker ensemble des données non organisées : traitement du langage naturel, reconnaissance automatique d'image ou de mouvement, « lacs de données », « data scraping » automatisé sur internet, ...

Les données numériques sont de plus en plus présentes dans notre environnement de travail quotidien

Pour nos usages professionnels liés à la connaissance : créer et récupérer de la donnée, faire des études et modèles, produire des cartes, penser des politiques ...

Pour produire des services nouveaux : partager de la donnée, « mieux » informer, permettre de réaliser une démarche en ligne, proposer des outils d'utilisation avancée de la donnée (téléchargement, cartographie...), permettre la création d'entreprises ...

Pour mieux interagir avec des usagers, citoyens, associations et experts qui vont enrichir les données : plateformes ouvertes, « Commons numériques », science participative ...

data.gouv.fr
Plateforme ouverte des données publiques françaises

Portail open data de l'ADEME :
132 jeux de données

Images satellitaires
Images LIDAR 3D de l'IGN

Observatoire des territoires
(ANCT)

transport.data.gouv.fr
Point d'Accès National aux données de transport

Vigicrues
Service d'information sur le risque de crues des principaux cours d'eau

Géoportail de l'urbanisme
Portail national de l'information réglementaire en urbanisme

EauFrance
Système d'information sur l'eau

http://www.toutsurlenvironnement.fr
Portail des informations environnementales des services publics

RenoIRH
Système d'information pour la gestion des agents

Service des données et des études statistiques (SDES)
Le service statistique ministériel

certificat-air.gouv.fr
Commander sa vignette Crit'Air

➔ Janvier 2023 : le ministère et son réseau recrutent
Géomaticien-nne - data science - mobilités (CEREMA)
Chargé.e de mission système d'information et innovation (DIRMED)
Enseignant.e-chercheur.se en science de la donnée massive en ingénierie transports et énergie (ENTPE)
Data Analyst H/F (ANAH)
Ingénieur.e data H/F (Observatoire de l'environnement en Bretagne)
Responsable des salles Serveurs et urbaniste du Data Center (F/H) (IGN)
Géomaticien.ne analyste (DDT Indre et Loire)
Spécialiste des systèmes d'information des certificats d'économies d'énergie (DGEC)
Chef-fe du bureau du droit des données personnelles et des technologies de l'information (DAJ)

inpn.mnhn.fr
Inventaire national du patrimoine naturel

demande-logement-social.gouv.fr
Demander un logement social et obtenir un numéro unique

ETC.

Plateformes régionales de données
DataSud, Data Grand Est, GéoPal, Open Data Occitanie, GéoNormandie, ...

En réalité la « donnée » numérique était là depuis quelques décennies
Mais on s'en rendait moins compte avant l'explosion des usages d'internet et la politique de mobilisation active de la donnée publique

Les **universités** de Grenoble, Toulouse et Paris ont été équipées d'ordinateurs IBM et ont commencé l'enseignement et la recherche en informatique en **1957**

L'**INSEE** (Institut national de la statistique et des études économiques) s'informatise depuis les **années 1960**

À la fin de la **décennie 1970**, la plupart des « **fonctions lourdes** » de **l'administration** sont informatisées : assiette et recouvrement des droits et taxes, paiement des dépenses des personnels, informatique statistique, modélisations diverses

Le **plan cadastral** français (tenu par les Finances Publiques) a commencé à être informatisé à la fin des **années 1980** (création de fichiers « images »)

En **1960**, la **SNCF** installe l'ordinateur français de BULL, le Gamma 60, à la gare d'Auteuil. La première application traitée sur cette machine est la solde. Elle faisait l'objet d'un traitement mécanisé depuis déjà 35 ans.

Les **Archives nationales** réalisent leurs premiers inventaires informatisés en **1973**.

Dès les **années 1980**, elles mènent une politique de conservation d'archives produites par les ministères sous format nativement numérique

Des milliers et des milliers de notes ont été créées par **traitement de texte** au sein de nos bureaux depuis le milieu des **années 1980** (où sont-elles aujourd'hui ?...)

Les premiers **systèmes d'information géographique** naissent dans les **années 1980** de la masse d'informations détenues par les collectivités territoriales

La loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés fonde le régime de protection de la donnée numérique personnelle
Elle crée la commission nationale de l'informatique et des libertés (CNIL)

Article 1 :

L'informatique doit être au service de chaque citoyen. Son développement doit s'opérer dans le cadre de la coopération internationale. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques.

Disposer de la donnée sous une forme numérique est tellement pratique qu'on numérise progressivement la donnée qui ne l'était pas à son origine

Les administrations de l'Etat et les données

Des missions :

- Utiliser les potentiels de la donnée pour adapter et améliorer le service public
- Produire de la donnée (statistique, recherche publique, Météo France, IGN, SHOM, OFB, etc.)
- Produire de la « donnée de référence » (9 jeux de données : Base Adresse Nationale, Plan Cadastral Informatisé, etc.)*
- Produire de la « donnée de forte valeur » *
- Réguler les usages du numérique

* Notions juridiques qui ne seront pas précisées ici

Des obligations juridiques :

- Communiquer certaines données sur demande
- Mettre en ligne par défaut certaines données et notamment
 - Les documents (dont les données) qui figurent dans le répertoire des principaux documents produits ou reçus par l'administration
 - Les bases de données, mises à jour de façon régulière, qu'elles produisent ou qu'elles reçoivent et qui ne font pas l'objet d'une diffusion publique par ailleurs ;
 - Les données, mises à jour de façon régulière, dont la publication présente un intérêt économique, social, sanitaire ou environnemental.

Cette mise en ligne doit s'opérer dans certaines conditions et notamment :

« Toute mise à disposition (...) sous forme électronique (...) se fait dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé »

- Rendre gratuite (sauf exception) la réutilisation des informations publiques (dont données et bases de données)
- Protéger les données personnelles incluses dans ce qui est communiqué et mis en ligne
- Sécuriser les secrets
- Partager des données entre administrations
- Appliquer des règles quand une décision individuelle est prise sur le fondement d'un traitement algorithmique
- Protéger le patrimoine de données sous forme d'Archives
- *A venir : restreindre certains usages de l'intelligence artificielle , ...*

Les administrations de l'Etat et les données

Une politique de la donnée publique (*aux sens stratégie, dynamique, organisation*)

Portée par :

- Le ministère de la transformation et de la fonction publiques
- La direction interministérielle du numérique (DINUM) et notamment son département Etalab
- L'administrateur général des données (actuellement la directrice de la DINUM)
- Des administrateurs ministériels des données (dans notre pôle ministériel ce sont le commissaire général au développement durable et le directeur de la stratégie et de l'innovation de l'agence nationale de la cohésion des territoires (ANCT))
- Des délégués ministériels à la protection des données personnelles (dans notre pôle ministériel c'est le directeur des affaires juridiques)
- Des hauts fonctionnaires chargés de la sécurité des systèmes d'information (dans notre pôle ministériel c'est le secrétaire général en sa qualité de Haut fonctionnaire de défense et de sécurité)

Précisée notamment par :

- La circulaire du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources
- Le cadre interministériel d'administration de la donnée du 23 septembre 2021
- Les feuilles de route ministérielles de la donnée, des algorithmes et des codes sources (septembre 2021)
- Les engagements de la France pris par exemple dans le cadre du « Partenariat mondial pour un gouvernement ouvert »

Un indispensable accompagnement technique

Infrastructures et réseaux, choix matériels, choix logiciels, articulation avec les prestataires, cybersécurité...

Fortement gérée ou influencée par :

- Des autorités administratives indépendantes (CNIL, CADA...), le Parlement, des instances de conseil (CESE, Conseil national du numérique, Conseil d'Etat, Cour des comptes, ...)
- La stratégie de l'UE pour la donnée et ses instruments

Une ambition

Que la France demeure à la 1^{ère} place dans les classements internationaux de qualité de la politique publique de la donnée

A. Le phénomène de numérisation de nos sociétés

La donnée préexiste au numérique. Mais le numérique renforce considérablement son potentiel

Le numérique a pris une place énorme dans notre organisation sociale grâce à ses avancées techniques (internet, stockage, vitesse, science algorithmique, objets connectés) et économiques (« big tech », taux d'utilisation d'internet par la population)

La donnée numérique peut être vue, au plus simple, comme la matière de base, le « sang » qui circule dans ses veines-réseaux

On assiste donc à la création d'une masse toujours plus considérable de données, sur laquelle les grands opérateurs numériques assoient leur richesse et leur potentiel (« *La donnée est le pétrole du 21^{ème} siècle* »)

Et les techniques de gestion de la donnée progressent en conséquence : moteurs de recherche, reconnaissance automatisée des images ou du langage, multiples usages de l'intelligence artificielle, stockages et services dans le Cloud, multiplication des interfaces automatisées ou API (interface de programmation d'application), ...

B. Les enjeux de la donnée pour les Etats / administrations

Les Etats / administrations ont aussi pris conscience du potentiel de la donnée, en visant leurs propres enjeux, qui sont multiples : mieux penser les politiques publiques, mieux informer le citoyen, mieux servir l'utilisateur du service public, faire des économies budgétaires, accroître la productivité, aider à la création de nouvelles entreprises, être compétitif en matière numérique au niveau international, protéger les libertés publiques, assurer la sécurité et la souveraineté, sauvegarder le patrimoine, promouvoir la recherche, etc.

La politique publique de la donnée (= des obligations, restrictions, incitations) vise ces multiples enjeux des Etats / administrations (y compris au niveau de l'Union européenne ou international) à travers ses dispositifs de : création, enrichissement, organisation, ouverture, partage, réutilisation, échange, protection, stockage, visualisation de la donnée publique et non publique

C. Les autres grands enjeux de la donnée pour notre pôle ministériel

Notre pôle ministériel, vu sa taille, génère et utilise (et doit ouvrir et protéger) **une grande masse de données** très utiles au fonctionnement quotidien de la Nation comme à la réussite de la transition écologique et énergétique (NB : lutter aussi contre les contre-vérités). C'est aussi le cas des collectivités, avec lesquelles notre pôle ministériel, et notamment les services déconcentrés, est en relation pour gérer le territoire. La **planification écologique** du projet France Nation Verte va avoir besoin de s'appuyer sur ces données existantes et à créer

Il existe des dispositifs juridiques spécifiques sur le droit de toute personne d'**accéder aux informations relatives à l'environnement**. Il existe aussi des dispositifs juridiques spécifiques sur le partage des données des transports, de l'énergie, des valeurs foncières, ...

Le pôle ministériel compte en son sein un important **réseau scientifique et technique** qui crée et publie des données de recherche

Le pôle ministériel doit former à la donnée dans ses **écoles** (ENTE, ENTPE, ENPC, ENSG, ENSM, ENSAM, ENAC, INSTN)

Plusieurs grandes politiques du pôle ministériel s'appuient sur des délégataires de service public, dont les **données** peuvent être qualifiés **d'intérêt général**

La directive du 14 mars 2007 établissant une infrastructure d'information géographique européenne (« **Inspire** ») a eu une très grande importance au sein de nos services pour structurer et ouvrir les jeux de données géographiques et environnementales. Dans les 228 données de références au sens d'Inspire, au moins 220 sont issues du pôle ministériel. Ceci explique que les géomaticiens fassent partie *de facto* du champ des métiers de la donnée (même si une montée en compétences est demandée)

Une grande partie des « **Données de forte valeur** » qui devront être partagées au niveau européen au 10 avril 2024 émaneront aussi de nos politiques publiques ministérielles : géospatiales ; observation de la Terre et environnement ; météorologiques ; mobilité (+ statistiques sur les entreprises et la population)

Plusieurs entreprises appartenant aux secteurs économiques régulés par le pôle ministériel font partie des « opérateurs de service essentiels » et des futurs « secteurs hautement critiques » des directives européennes sur la **cybersécurité** de 2016 et 2022 : transports, énergie, eaux potables et usées

Le pôle ministériel a en charge des enjeux de **régulation** d'usages experts de la donnée : drones, véhicules autonomes, ville « intelligente » (*smart city*), ...

Le saviez-vous ?

- L'agence nationale de la cohésion des territoires (**ANCT**), opérateur de notre pôle ministériel depuis juillet 2022, porte les politiques d'accès des territoires et des personnes au numérique et donc à de multiples usages de la donnée
- **Météo France**, opérateur de notre pôle ministériel, possède deux des plus gros supercalculateurs français pour faire tourner ses modèles : Belenos et Taranis
- 72 heures après son ouverture en 2006, le Géoportail de l'**IGN** avait reçu 13 millions de connexions (et était devenu inaccessible)

Pour aller plus loin...



1er juin 2023

Journée de la donnée
du pôle ministériel



Économie et gouvernance
de la donnée
Soraya Duboc et Daniel-Julien Noël

Feuille de route de la donnée, des algorithmes et des codes sources
Le Mardi 17 janvier 2023

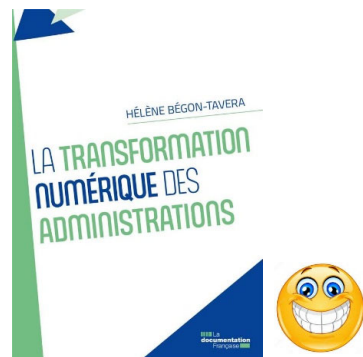
Pilotée par le Commissariat général au développement durable, la feuille de route propose un cap pour améliorer l'usage de la donnée publique et accélérer la transition écologique sur un horizon de trois à quatre ans. Elle constitue une réponse à la nouvelle politique publique de la donnée inscrite dans la circulaire du Premier ministre du 27 avril 2021 relative à une politique de la donnée, des algorithmes et des codes sources.



Etalab
Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

Politique publique de la donnée

**Ouvrir, partager et valoriser
les données publiques**



RÉPUBLIQUE FRANÇAISE
Liberté • Égalité • Fraternité

AVIS DU CONSEIL NATIONAL DU NUMÉRIQUE
JUILLET 2020

**FAIRE DES DONNÉES ENVIRONNEMENTALES
DES DONNÉES D'INTÉRÊT GÉNÉRAL**

**Pour une politique publique
de la donnée**
Rapport

Eric Bothorel - Stéphanie Combes –
Renaud Vedel, décembre 2020