

OpenData

# La Data, c'est quoi ?

## CEREMA Med > Mathieu Rajerison

02/08/2018

# La Data

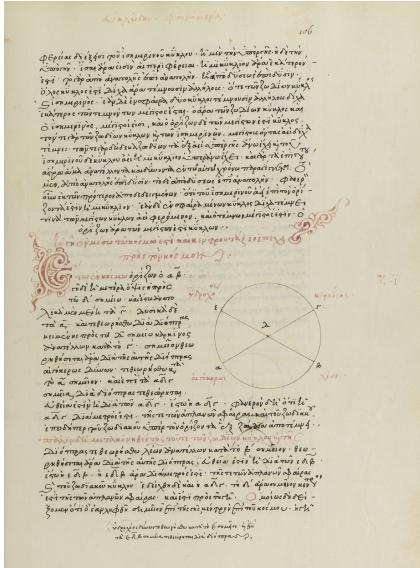
# Au quotidien

Notre quotidien est baigné par les données.

- Dans la sphère *intime*, notre consommation énergétique est transmise à notre fournisseur d'électricité, nos revenus sont déclarés sur des feuilles d'imposition. Nous donnons des informations personnelles via les mails que nous échangeons, recevons. Des données transitent au sujet des achats que nous réalisons sur Internet et sur le montant et la nature de nos transactions bancaires.
- Dans la sphère *professionnelle*, nous échangeons des courriels, exploitons des documents, des tableaux, remplissons des formulaires sur applications web métier.

De quelle langue vient data ?

# Étymologie, racines



“

L'un des premiers usages scientifiques connus du terme « data » (issu du verbe latin « dare » : donner) remonte au IIIe siècle avant J.-C. dans un traité de mathématiques d'Euclide où il était démontré des propriétés géométriques « étant donné » d'autres propriétés géométriques - autrement dit des axiomes. Son sens a ensuite évolué en deux branches. D'une part, il a donné le mot « date », que l'on retrouve dans des documents médiévaux officiels pour signifier le « moment » où ceux-ci ont été produits ; d'autre part, à compter du XVIIe siècle, le sens de « données » en référence à des « principes acceptés de tous », par l'intermédiaire du divin en théologie, puis par l'expérimentation en mathématiques.

Histoire des sciences de la Data

On retrouve dans la définition des data un caractère "**gratuit**" et immédiat, ainsi que d'objectivité et de **factualité**.

# A quoi ressemble une donnée ?

Une spécialiste de la donnée en donne une description assez originale :

“

Il s'agit d'une masse abstraite de nombres, sans poids et non identifiée (la plupart du temps encodée numériquement), ayant une influence puissante sur nos vies.

Concrètement, ces données nous apparaissent structurées au travers de tableaux, de tables de données.

## Voici quelques exemples visuels de données :

indicateurs-dd-2017-épanouissement-des-êtres-humains.xls - LibreOffice Calc

A	B	C	D
7	Code géographique du territoire (*)	Libellé géographique du territoire (*)	2013
8	200000628	CC Rhône Lez Provence	21,1
9	200004802	CC du FM de Fayence	15,6
10	200027100	CC Méditerranée Porte des Maures	13,5
11	200030195	Métropole Nice Côte d'Azur	17,3
12	200034700	CA Durance-Luberon-Verdon Agglo	15,4
13	200035087	CA Terre de Provence	17,4
14	200035319	CA Var Estrel Méditerranée (Cavem)	15,1
15	200035723	CC Ventoux Sud	21,5
16	200036077	CC du Golfe de Saint-Tropez	15,9
17	200039857	CA du FM de Grasse	12,8
18	200039915	CA Cannes FM de Lérins	16,7
19	200039931	CC Alpes d'Azur	18,7
20	200040202	CC Provence Verdon	16,3
21	200040210	CC Lacs et Gorges du Verdon	20,4
22	200040442	CA Luberon Monts de Vaucluse	20,6
23	200040624	CC FM d'Apt-Luberon	19,9
24	200040681	CC Enclave des Papes-FM de Grignan	18,9
25	200054807	Métropole d'Aix-Marseille-Provence	18,3
26	200060320	CC Serre-Ponçon Val d'Avance	12,9
27	200067437	CA Provence-Alpes-Côte d'Azur	15,6
28	200067445	CC Béch-Bélaly	14,3
29	200067452	CC du Guillestrois et du Queyras	13,0
30	200067472	CC Serre-Ponçon	13,6
31	200067825	CA Gap-Tallard-Durance	14,1
32	200068096	CC Champsaur-Valgaudemar	15,1
33	200068101	CA de la Drôme provençale	14,4

env.ini | xwingen.html | config.R | indicateurs-dd-2017-épanouissement-des-êtres-humains.csv

1	Code géographique du territoire (*)	Libellé géographique du territoire (*)	2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,
2	9301,Digne-les-Bains,"7,5°","7°,7°","8,0°","8,3°","8,4°","8,2°","9,1°","9,6°","10,5°","11,2°","11,5°","11,7°",		
3	9302,Manosque,"9,9°","9,9°","9,9°","10,1°","9,9°","9,2°","8,8°","10,5°","10,6°","10,9°","11,8°","12,4°","12,3°","12,4°",		
4	9303,Briançon,"7,2°","7,5°","7,7°","7,9°","7,4°","6,5°","7,3°","7,5°","7,5°","8,1°","8,9°","9,1°","9,1°",		
5	9304,Gap,"7,0°","7,2°","7,4°","7,2°","6,9°","6,4°","7,3°","7,5°","8,0°","8,5°","9,0°","8,9°","9,1°",		
6	9305,Cannes-Antibes,"9,5°","9,4°","9,3°","9,0°","8,1°","7,1°","7,8°","9,2°","9,3°","9,9°","10,5°","10,6°","10,7°",		
7	9306,Menton Vallée de la Roya,"6,8°","6,8°","6,7°","6,9°","5,9°","5,3°","6,9°","7,2°","7,6°","8,5°","8,2°","9,5°","9,8°",		
8	9307,Nice,"8,1°","8,1°","7,9°","7,2°","6,6°","8,2°","8,3°","8,5°","9,3°","10,1°","10,5°","10,7°",		
9	9308,Aix-en-Provence,"9,3°","9,0°","8,8°","8,1°","7,2°","8,5°","8,6°","8,8°","9,1°","9,5°","9,5°","9,6°",		
10	9309,Arles,"11,3°","11,2°","11,3°","11,9°","11,1°","10,3°","12,4°","12,9°","13,5°","14,2°","14,7°","14,5°","14,2°",		
11	9310,Marseille-Aubagne,"11,6°","11,5°","11,5°","11,8°","11,2°","10,1°","11,5°","11,9°","12,1°","12,4°","12,7°","12,6°","12,7°",		
12	9311,Istres-Martigues,"10,4°","10,2°","10,1°","10,1°","9,3°","8,7°","10,8°","11,1°","10,8°","10,9°","11,4°","11,2°","11,2°",		
13	9312,Salon-de-Provence,"10,6°","10,4°","10,1°","10,0°","9,7°","9,0°","10,8°","11,1°","11,4°","11,7°","12,1°","12,1°","12,2°",		
14	9313,Draguignan,"11,3°","11,2°","11,0°","10,6°","10,2°","9,6°","11,2°","11,3°","11,4°","11,9°","12,6°","12,8°","12,7°",		
15	9314,Fréjus-Saint Raphaël,"10,2°","11,1°","10,9°","10,5°","9,8°","8,9°","10,7°","10,9°","11,2°","11,7°","12,2°","12,4°","12,6°",		
16	9315,Toulon,"10,5°","10,5°","10,3°","9,9°","9,1°","8,4°","9,7°","9,9°","9,9°","10,3°","10,9°","10,9°",		
17	9316,Cavaillon-Apt,"9,1°","9,1°","9,5°","9,7°","9,1°","8,3°","10,5°","10,7°","11,3°","12,0°","12,6°","12,7°","13,1°",		
18	9317,Orange,"9,4°","9,5°","10,2°","10,5°","10,1°","9,3°","11,2°","11,4°","11,6°","12,5°","13,1°","13,4°","13,6°",		
19	9359,Avignon – partie Provence-Alpes-Côte d'Azur,"9,1°","9,3°","9,6°","9,7°","9,1°","8,4°","10,1°","10,6°","11,1°","11,7°","12,2°","12,2°","12,3°",		

Édition des données - serveur dvf (172.23.210.147:5432) - mnca\_v2 - result:t2\_dta

	periode_text	nom_dta	nb_appt	nb_maison	nb_dependance	nb_loc_activite	nb_tab	nb_autre_terrain	pk	nom	character varying(90)	typ	character varying(20)	the_geom
1	2014 - 2016	dta_littoral	26164	1685	4592	2261	157	271	50	dta_littoral	dta			
2	2014 - 2016	dta_moyen_pays	1647	1269	252	289	302	360	51	dta_moyen_pays	dta			
3	2014 - 2016	dta_haut_pays	788	288	122	307	43	378	949	dta_haut_pays	dta			

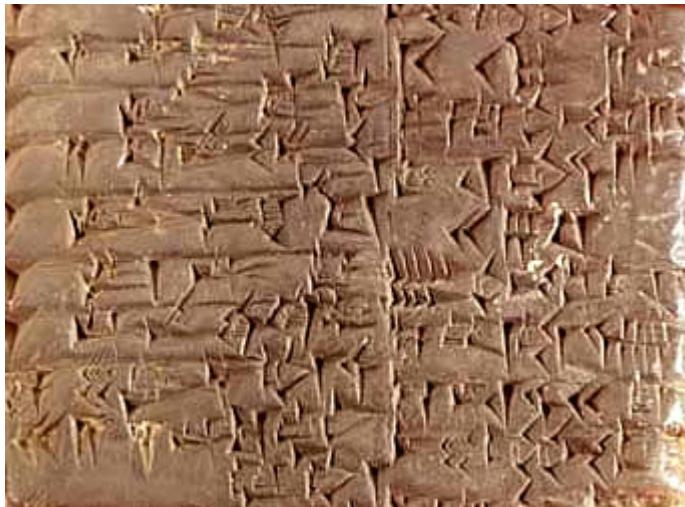
Citez trois modes de communication

# Naissance de la data

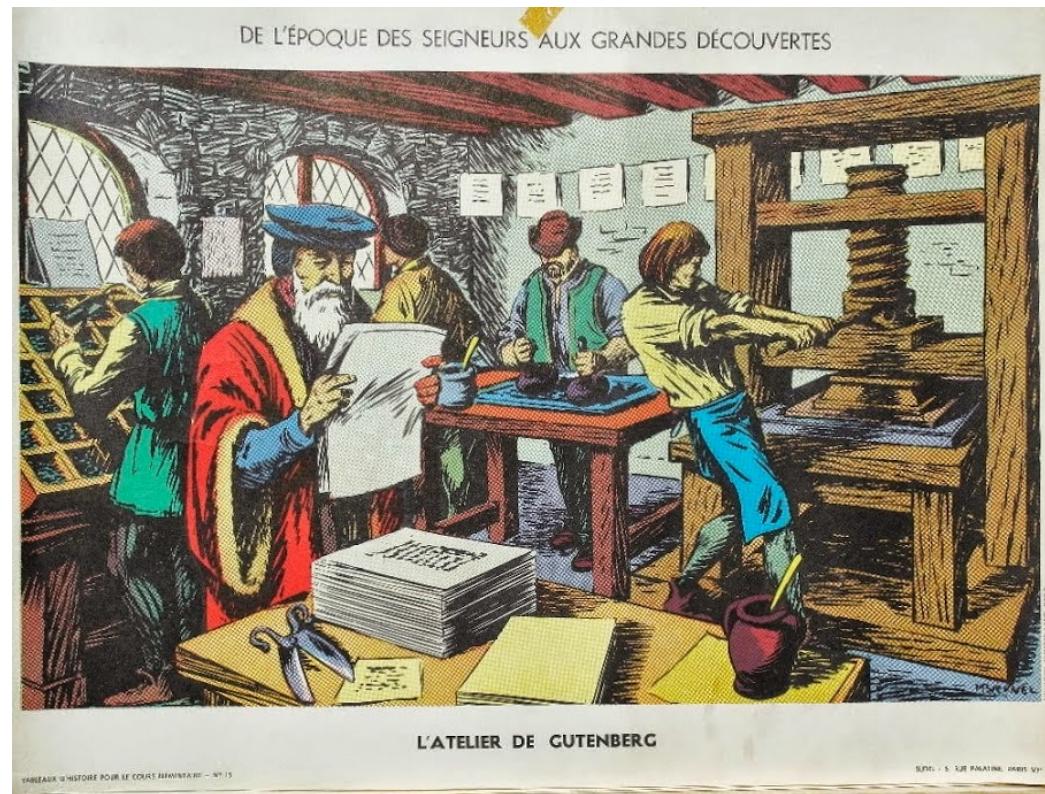
- En premier lieu était **la transmission orale** des connaissances



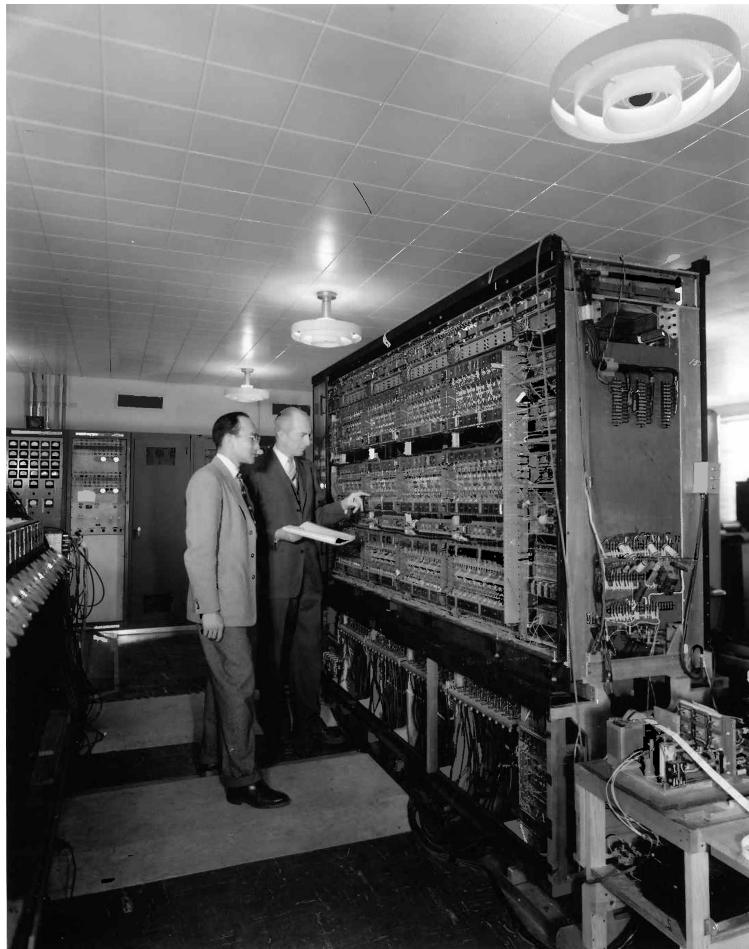
- Puis vint l'**écriture**



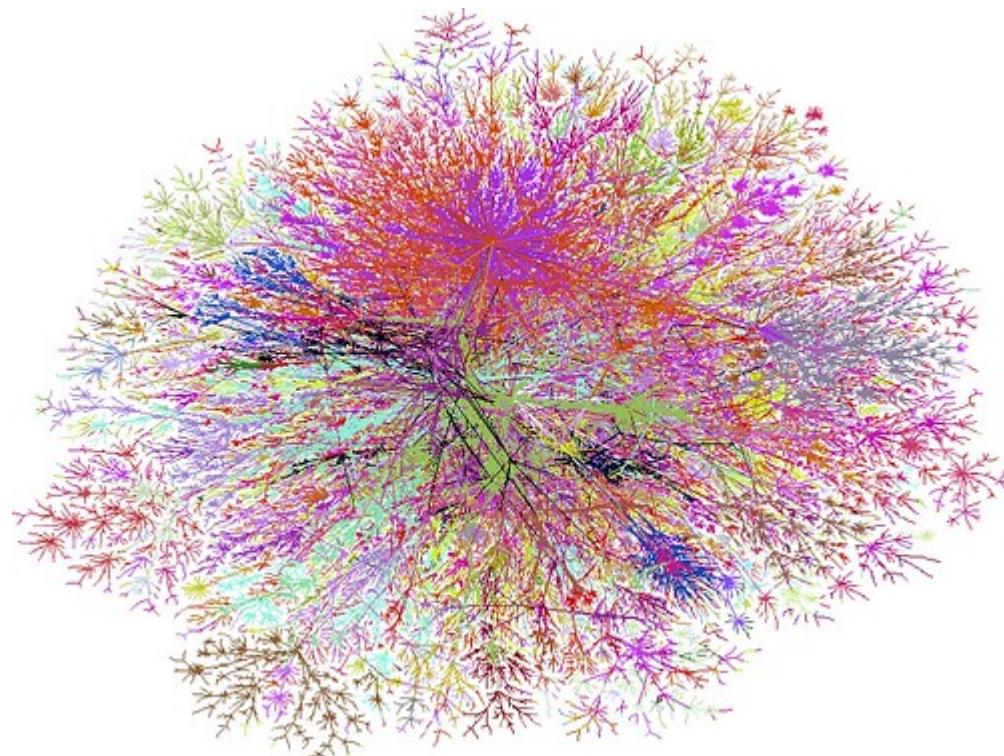
- L'imprimerie, née avec Gutenberg, qui permit de reproduire et, ainsi, de favoriser la diffusion des écrits



- Le **numérique** a permis, quant à lui, de confier le stockage et l'interprétation des savoirs à des machines.



- **Internet** et la corrélation des savoirs, qu'illustre le web (*toile* en français), sont nées de cette révolution technologique.



# Le Big Data

Dans quoi sont stockées les données ?

Par quelle voie transitent-elles ?

Quel élément de l'ordinateur permet de faire des calculs ?

Le courant du Big Data suit celui de l'accroissement :

- des *flux*
- des capacités de *stockage* des données
- des capacités de *calcul* des ordinateurs

# Flux

L'accroissement des flux de données concerne par exemple celui des réseaux de téléphonie mobile, de l'internet, qui permettent tout un tas d'applications telles que la télémédecine, l'information en temps réel pour les agriculteurs, d'autres modalités de travail (télétravail),....

Les données circulent entre des machines soit via les airs, en particulier via les ondes via les satellites et antennes émettrices, soit via des câbles comme la fibre optique.



# Stockage

L'augmentation du stockage des données est une des plus grandes caractéristiques du big data. Celle-ci permettra de dématérialiser de plus en plus de données, et en corollaire des procédures, telles que les déclarations d'impôt.

Les données sont stockées sur des supports tels que les disques durs, des clés USB à usage souvent personnel. Même logées dans le cloud, les données sont sur ces supports.



# Calcul

Les ordinateurs sont dotés de capacités de calcul allant grandissant (loi de Moore).

Les calculs, sur un ordinateur, sont réalisés grâce aux **processeurs**, parfois aussi grâce à la carte graphique, sur la base d'une binarisation de l'information. Ceci, en attendant l'arrivée de l'ordinateur quantique qui considérera des états intermédiaires (ou plutôt simultanés) entre le 0 et 1.

Quelqu'un peut-il me dire comment il calcule  
 $256 + 125$  ?

Les **algorithmes** sont des séquences ordonnées d'opérations informatiques. Ils automatisent le traitement de masses d'information et parfois en révèlent des motifs. Certains de ces algorithmes appelés réseaux de neurones miment même le fonctionnement du cerveau humain et sont en plein développement (intelligence artificielle).



Quelles sont les différentes étapes de la "vie" d'un produit de consommation ?

# L'écosystème de la donnée

## La chaîne des données

La chaîne des données se compose des étapes suivantes :

- produire
- mettre en circulation
- ouvrir
- réutiliser
- gouverner

# Les acteurs

Dans l'écosystème de la donnée, on distingue plusieurs acteurs :

- les **producteurs** de données qui génèrent des données, de façon active ou pas (capteurs)
- les **collecteurs** qui rassemblent les données émises sur le web
- les **administrateurs** qui classent et documentent la donnée collectée
- les **réutilisateurs** qui exploitent, "font parler" les données
- les **contributors** qui, le plus souvent de leur propre fait, participent à l'alimentation d'une base de données (comme wikipedia)

# La Data, liant d'un monde connecté

Citez quelques objets connectés

# Les objets connectés

L'internet des objets concerne des objets du quotidien capables de transmettre, d'émettre et de réagir à des informations circulant sur le réseau internet.

Le premier usage est **domotique**, pour la maison avec le pilotage à distance de la climatisation d'un logement, la surveillance...

Il peut aussi concerner des **véhicules**, des réfrigérateurs, des objets récréatifs tels que des enceintes...

...voire même des poupées



Article de The Guardian : "German parents told to destroy doll that can spy on children"

# La Smart City ou ville intelligente

La Smart City, ou ville intelligente, est une ville dans laquelle de multiples **capteurs**, voire des **données issues des téléphones des citoyens**, remontent à un siège central des informations sur la ville : congestion routière, pollution de l'air, consommation énergétique, ...

On lui appose le préfixe *smart* car elle est censée s'adosser, pour diriger ses opérations, sur la **synthèse de données** factuelles variées, présentes en très grand nombre.

L'objet de la Smart City est de créer une cité harmonieuse quasiment autogérée par les algorithmes et les données captées et émises par le réseau.



Pour certains, Smart City rime aussi avec *Big Brother*, notamment vidéosurveillance et sécurité publique. Les images peuvent être interprétées par des ordinateurs afin d'identifier des attroupements mais aussi des individus.

» Article de la Quadrature du Net : "La Smart City policière se répand comme une traînée de poudre"

» Reconnaissance faciale en Chine, réalisée par une startup de Shanghai :



# Les données et le citoyen

A votre avis, quelles données transmettez-vous, parfois à votre insu ?

# L'individu base de données

Un citoyen connecté à internet alimente, parfois sans le savoir, les bases de données de sociétés ayant monopole sur le réseau. Les données qu'il émet peuvent concerner :

- Tout d'abord, ses **envies**, ses désirs, ses besoins et ses interrogations, par le biais d'un moteur de recherche tel que Google.
- Les **informations personnelles** partagées sur un réseau social comme Facebook ou bien Gmail.
- Sa **localisation**, s'il a activé le GPS de son smartphone
- Ses **goûts**, notamment musicaux, s'il se connecte à Spotify ou a commandé des disques sur Amazon.

Le citoyen est tellement connecté et parfois si mal informé que des données personnelles échappent à sa vigilance et sont transmises à des tiers. Les conséquences en cascade de ces fuites peuvent largement s'étendre hors de sa sphère intime.

En l'occurrence, peu de personnes lisent les *Conditions Générales d'Utilisation* d'applications Internet avant de les approuver.

# Impact des données

## Pour ceux qui s'intéressent aux citoyens

L'essor d'internet et de ses grands groupes est intimement lié à l'accès aux données personnelles des individus, d'où le fait de qualifier les données comme de l'*or noir du XXI<sup>e</sup> siècle*.

Les GAFA et BATX en tirent bénéfice en **personnalisant** les pubs que nous recevons et en recueillant des informations au sujet de nos envies et préférences.

**Les données individuelles sont précieuses pour les villes intelligentes.** En effet, la cohérence des décisions prises par l'ordinateur central dépendra de la qualité, de la fraîcheur et de la fréquence des données recueillies auprès des individus eux-mêmes ou de l'environnement plus ou moins proche de ces derniers : immeuble, bloc, quartier, ville,...

A votre avis, quel avantage peut avoir un gouvernement ou une société à ouvrir ses données ?

## Pour les citoyens qui s'intéressent à la société ou à leur environnement

Enfin, la livraison de données est un des éléments de la démocratie où les citoyens ont le **droit d'avoir accès** à certaines décisions publiques, aux budgets.

De plus en plus, l'ouverture des données confère aux pays, voire, aux collectivités qui la mettent en oeuvre, un caractère d'**exemplarité**.

Nous le verrons plus tard : les données, lorsqu'elles sont ouvertes, permettent même de voir fleurir des entreprises qui bâtissent des services basés sur leur exploitation. **L'ouverture des données dynamise l'économie**.

# Protection des données

## CNIL

La Commission Nationale de l'Informatique et des Libertés est une autorité administrative française chargée de veiller à ce que l'informatique soit au service du citoyen et **qu'elle ne porte atteinte ni à l'identité humaine, ni aux droits de l'Homme, ni à la vie privée, ni aux libertés individuelles ou publiques.**

## RGPD

Face au commerce lucratif des **données personnelles** réalisé parfois à l'insu des utilisateurs et la sensibilité de certaines données à caractère personnel, l'Union Européenne s'est dotée cette année du Règlement Général sur la Protection des données (RGPD)

# La culture des données ou Data Literacy

Ce n'est pas parce qu'une donnée est en libre accès qu'elle est accessible.

L'accès et l'exploitation des données est **souvent l'affaire de spécialistes de la donnée** alors que le contenu de celles-ci est censé intéresser tout le monde.

Afin de se départir d'une vision élitiste de la donnée, la Data Literacy (Littératie des données, ou plus simplement la culture de la donnée), a pour but d'amener à s'intéresser aux données des personnes non-spécialistes.

Cela passe par la **formation**, la **sensibilisation** des individus.

→ **Testez votre culture des données avec dataposition**