



devient



Data et IA en pleine confiance

# L'éthique comme condition du développement des Big Data

Atelier de réflexion éthique Genotoul

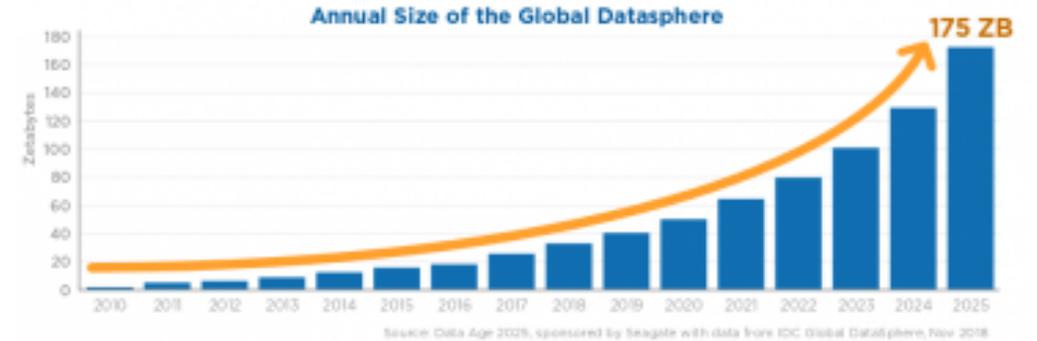
20 mai 2021



*Pour toute utilisation du contenu de cette présentation, veuillez citer l'auteur, son organisme d'appartenance, le titre et la date du document, ainsi que le volet 3 de l'atelier 2020-2021 « Usages éthiques des Big Data en biosciences » de la Plateforme Ethique et Biosciences (Genotoul Societal) de Toulouse. Merci.*

# De l'effet Woaw...

- Explosion des données produites
- Explosion des capacités de traitement logiciel
- Explosion des usages du Big data et de l'IA : reconnaissance d'images, réseaux sociaux, santé connectée, publicité ciblée...



### The ImageNet dataset [Fei-Fei et al. 2012]

- ▶ 1.2 million training samples
- ▶ 1000 categories

### Fast & Programmable General-Purpose GPUs

- ▶ NVIDIA CUDA
- ▶ Capable of over 1 trillion operations/second



Matchstick



Sea lion



Flute



Strawberry



Backpack



Bathing cap



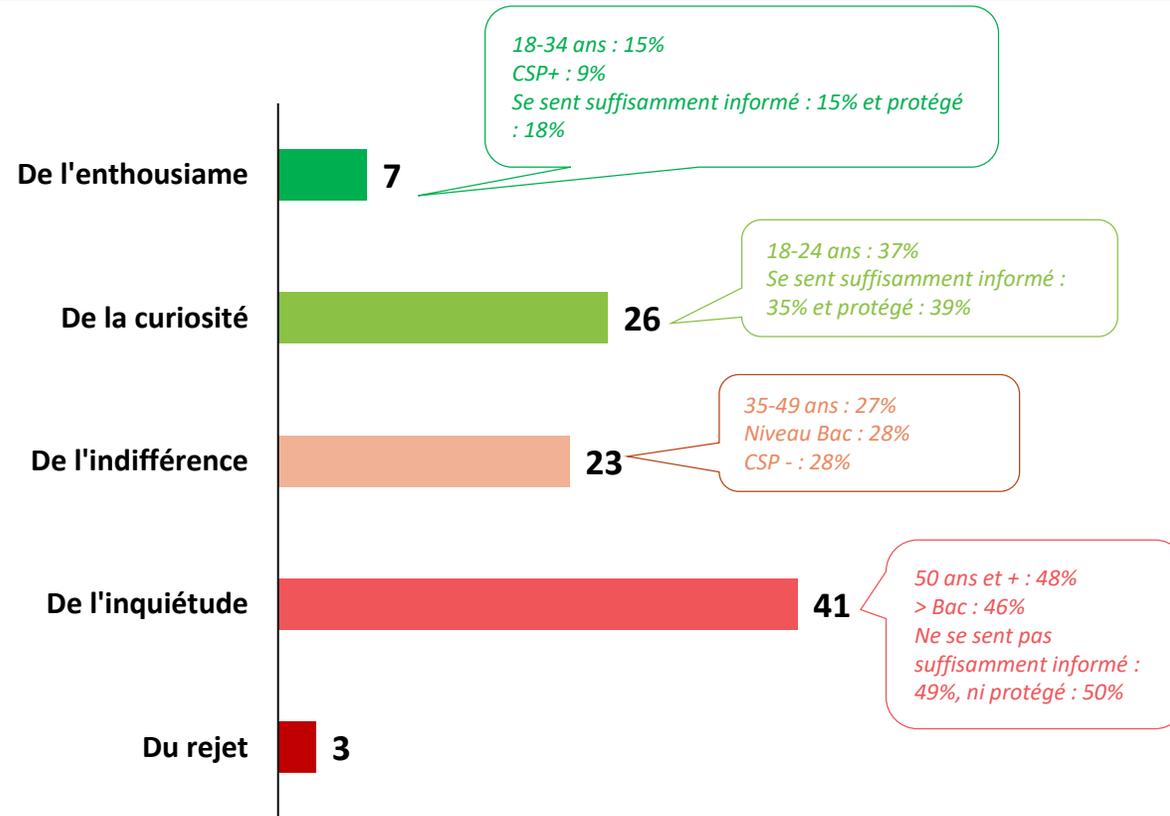
Racket



## à l'effet Hum...



Lorsque vous entendez parler du sujet des data, c'est-à-dire des données numériques dans leur ensemble, qu'elles soient personnelles ou collectives, qui peuvent circuler par un réseau informatique, et qui sont collectées, stockées et utilisées par les moteurs de recherche, par les réseaux sociaux, par les entreprises dont vous êtes client : que ressentez-vous principalement ... ?



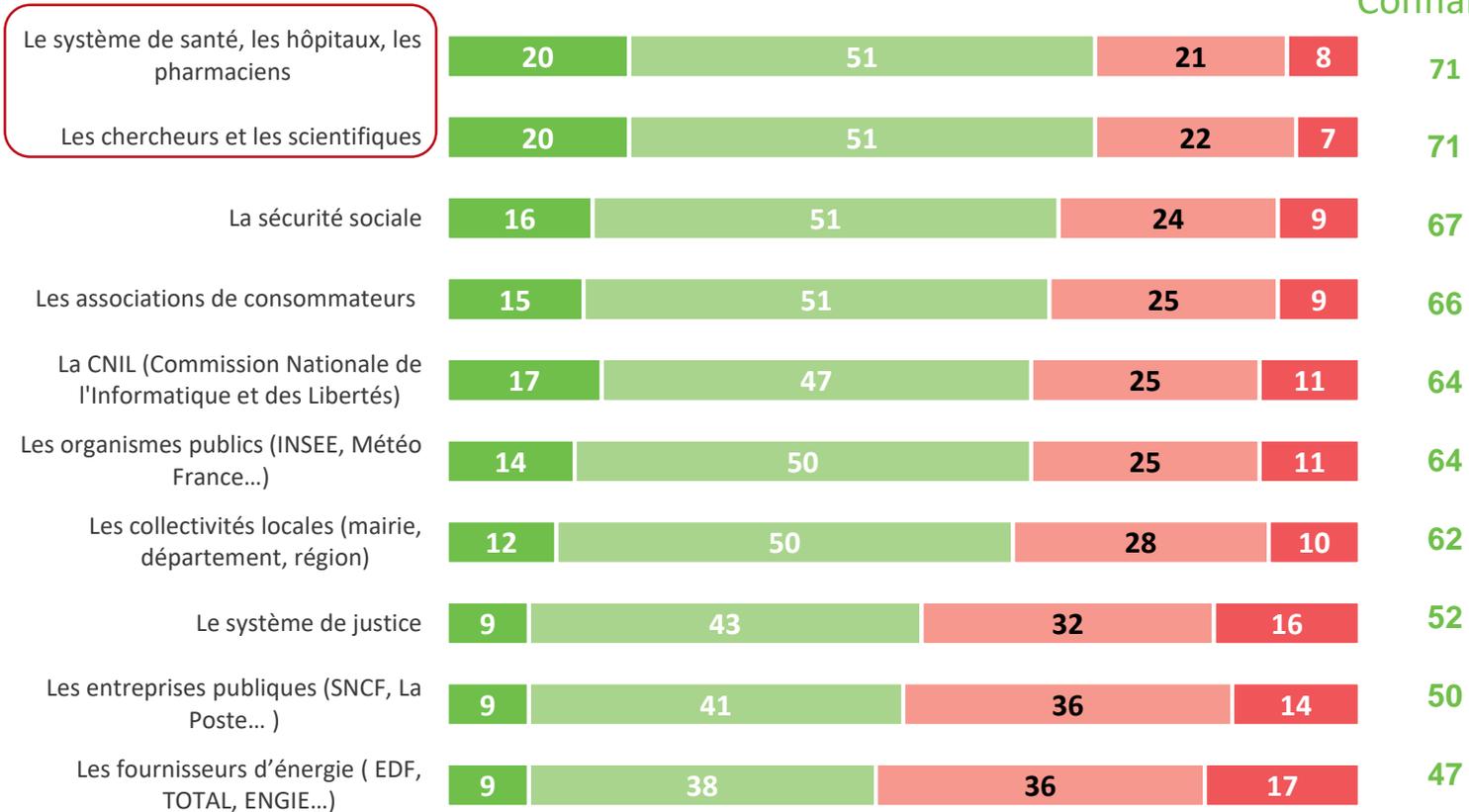
Source : sondage Harris Interactive et CAP pour Occitanie Data et CIVITEO, septembre 2019

Pour chacun des acteurs suivants, pouvez-vous me dire si concernant la gestion et l'utilisation des données en France vous lui faites tout à fait confiance, assez confiance, assez peu confiance ou pas confiance du tout ?

- À tous - en %

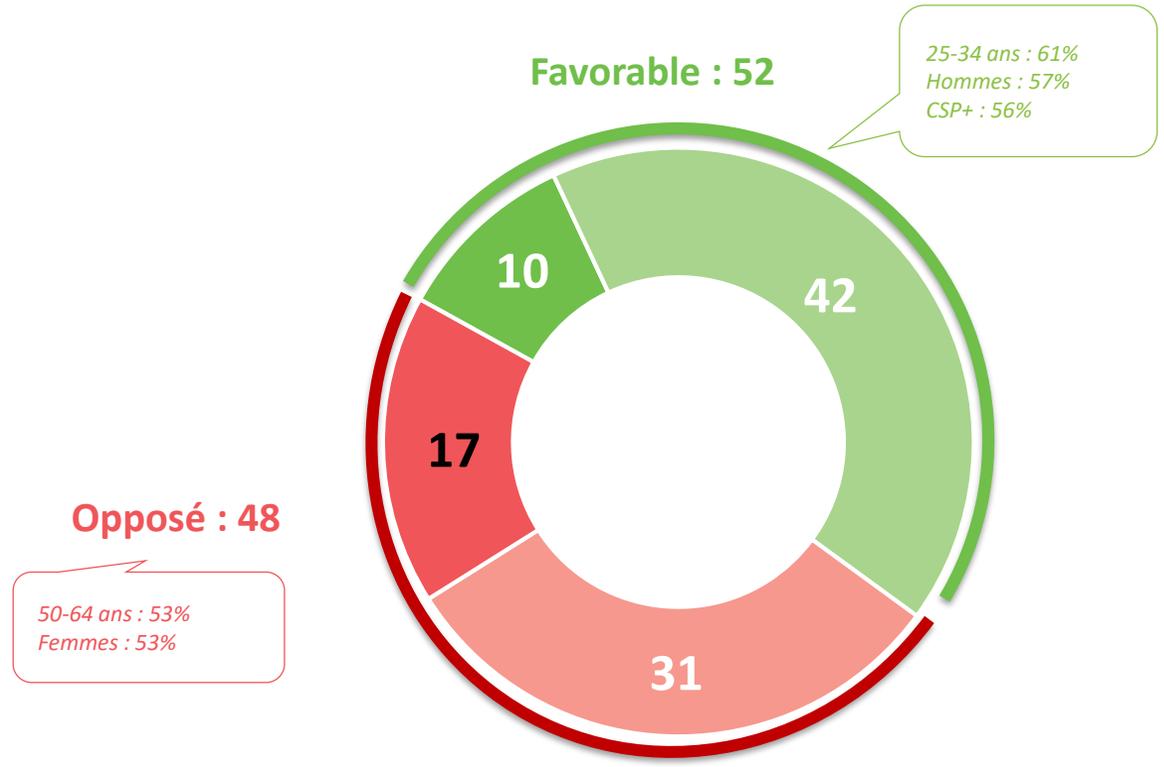
1/2

Total  
Confiance



■ Tout à fait confiance ■ Assez confiance ■ Assez peu confiance ■ Pas du tout confiance

Etes-vous tout à fait favorable, assez favorable, assez opposé ou très opposé à ce que des acteurs publics et privés puissent échanger des données et les exploiter afin d'améliorer des services publics, services d'intérêt général, produits et services privés ?  
Cela se ferait sous certaines conditions d'engagements et de contrôles précis au-delà des réglementations actuelles nationales contrôlées par la CNIL.

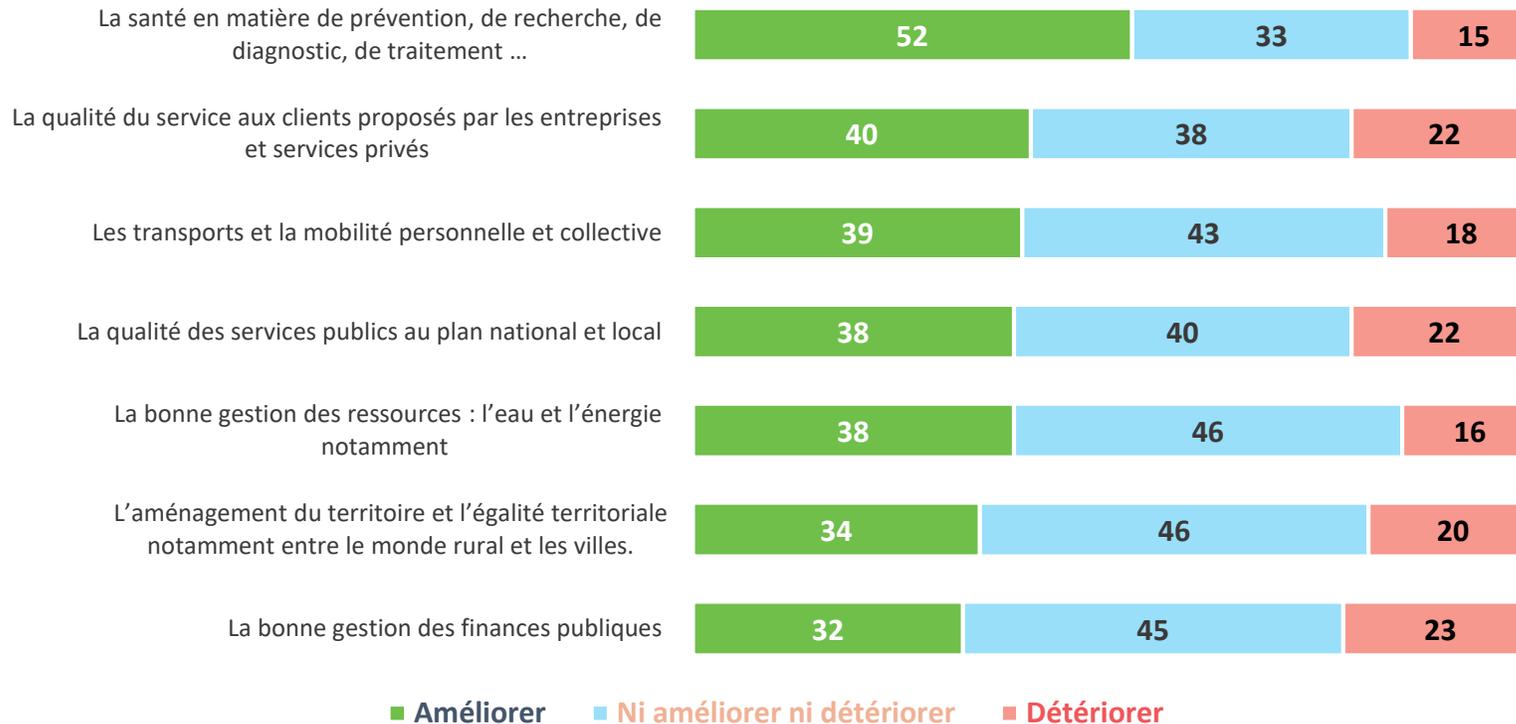


■ Tout à fait favorable ■ Assez favorable ■ Assez opposé ■ Tout à fait opposé

**Plus précisément, pensez-vous que l'usage des données par des acteurs publics et privés soit susceptible d'améliorer, de détériorer ou, ni d'améliorer ni de détériorer, chacun des points suivants ?**

- À tous - en %

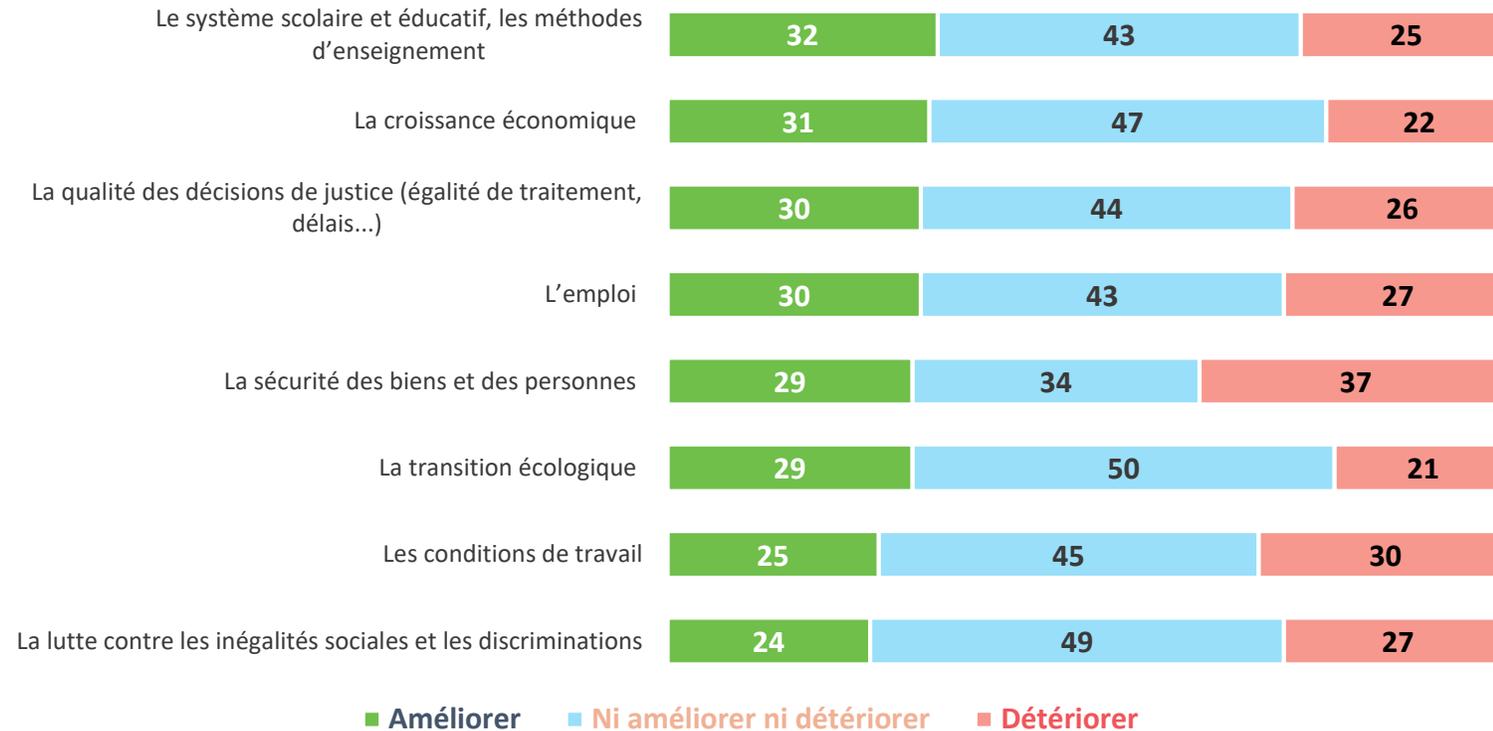
1/2



**Plus précisément, pensez-vous que l'usage des données par des acteurs publics et privés soit susceptible d'améliorer, de détériorer ou, ni d'améliorer ni de détériorer, chacun des points suivants ?**

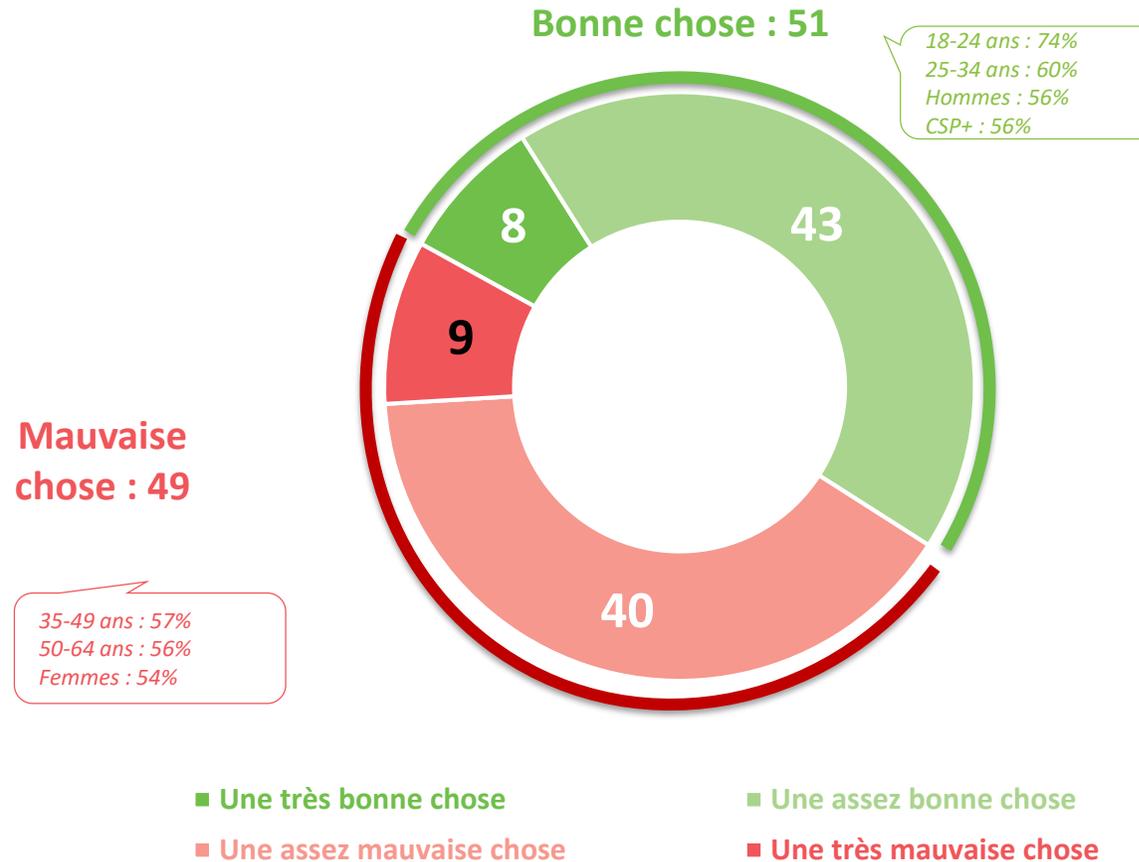
2/2

- À tous - en %



**Au total, pensez-vous que l'usage croissant des données constitue une très bonne chose, une assez bonne chose, une assez mauvaise chose ou une très mauvaise chose pour l'évolution de la société ?**

- À tous - en %



slido

Rejoindre sur  
**slido.com**  
**#Ethique**



# L'accueil du projet Occitanie Data

Si votre région prenait cette initiative estimez-vous qu'il s'agirait d'une très bonne initiative, assez bonne initiative, assez mauvaise initiative ou très mauvaise initiative ?

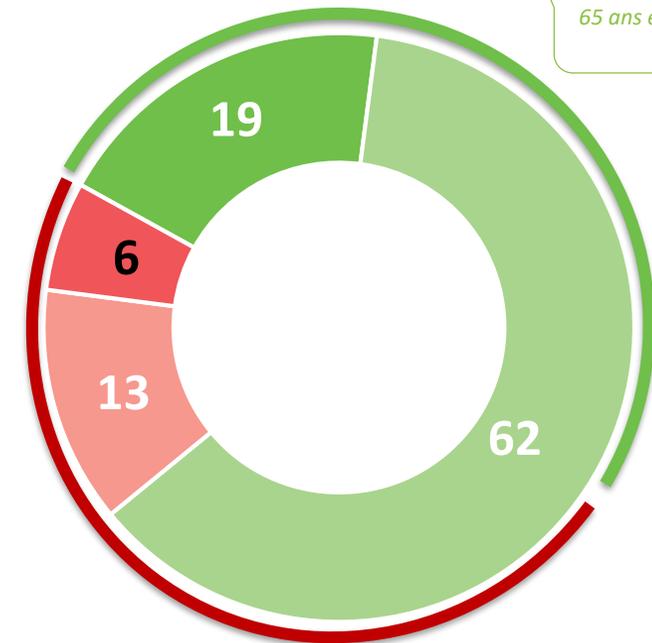
Un conseil régional rassemble des partenaires publics et privés pour construire une « plateforme numérique », permettant d'échanger et d'exploiter de manière sécurisée des données pour offrir de nouveaux services d'intérêt public aux citoyens : par exemple pour mieux maîtriser la dépense énergétique dans la région, offrir une meilleure couverture haut débit, mieux prévenir les catastrophes naturelles (inondations, sécheresse...), optimiser les horaires de passage des transports collectifs ...

Mauvaise initiative : 19

35-49 ans : 24%  
CSP- : 24%

Bonne initiative : 81

65 ans et + : 87%



■ Une très bonne initiative

■ Une assez bonne initiative

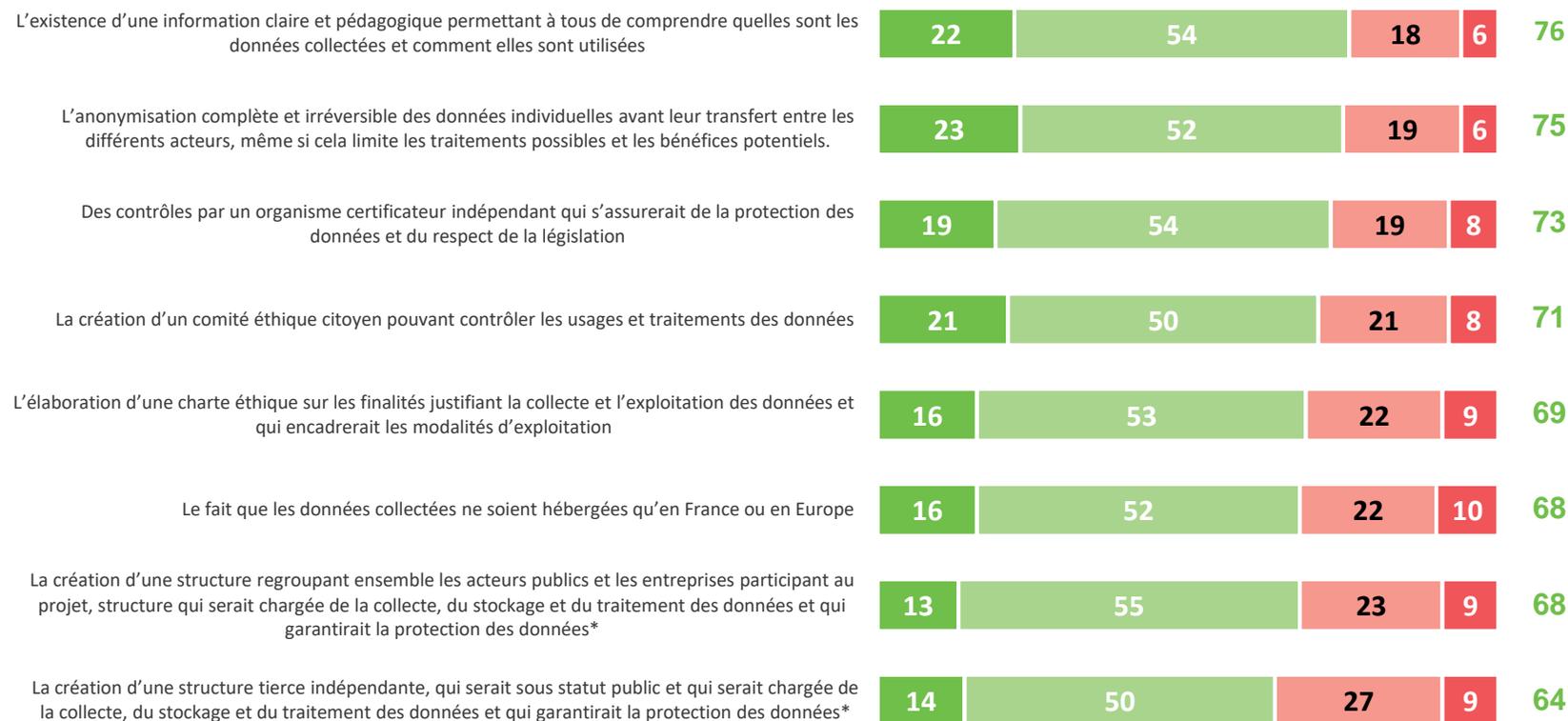
■ Une assez mauvaise initiative

■ Une très mauvaise initiative

# Pourquoi ?

Quand vous pensez à ce type d'initiative, pouvez-vous dire si chacun des éléments suivants serait ou pas de nature à vous donner confiance dans une telle démarche ?

**Total  
Confiance**



■ Tout à fait confiance

■ Assez confiance

■ Assez peu confiance

■ Pas du tout confiance

\*(les données ne transiteraient pas entre chacun des acteurs)

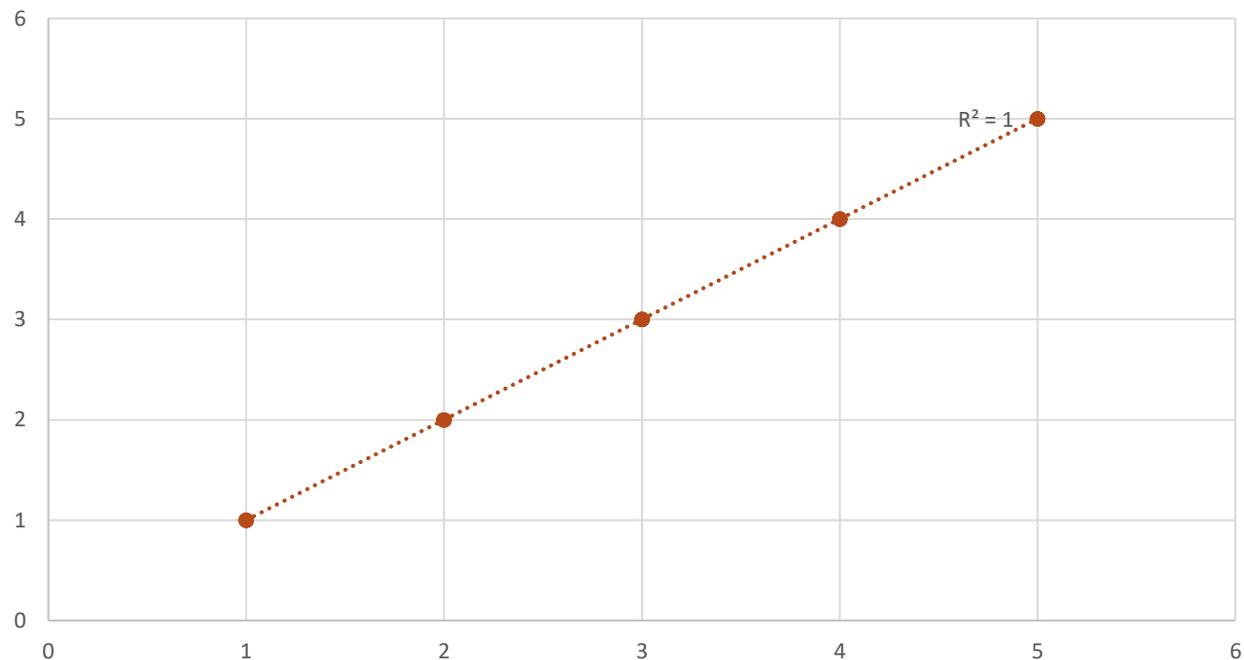
# Les limites de l'anonymisation

Conférence « Quand les maths éclairent l'éthique », avec Sébastien Lamy, chercheur INSERM, UMR 1027 UPS-INSERM, mars 2020

# UN EXEMPLE SIMPLIFIÉ

Id	V1	V2	Loc
1	1	1	1
2	1	1	2
3	1	1	3
4	1	1	4
5	2	2	1
6	2	2	2
7	2	2	3
8	2	2	4
9	3	3	1
10	3	3	2
11	3	3	3
12	3	3	4
13	4	4	1
14	4	4	2
15	4	4	3
16	4	4	4
17	5	5	1
18	5	5	2
19	5	5	3
20	5	5	4

Valeurs réelles

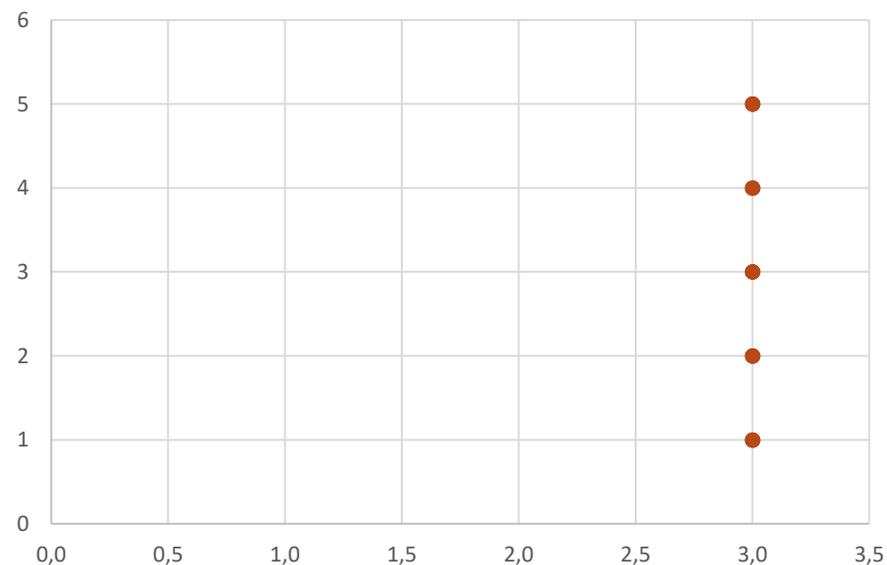


# UN EXEMPLE SIMPLIFIÉ

Id	V1	V1_i	V2	Loc
1	1	3	1	1
2	1	3	1	2
3	1	3	1	3
4	1	3	1	4
5	2	3	2	1
6	2	3	2	2
7	2	3	2	3
8	2	3	2	4
9	3	3	3	1
10	3	3	3	2
11	3	3	3	3
12	3	3	3	4
13	4	3	4	1
14	4	3	4	2
15	4	3	4	3
16	4	3	4	4
17	5	3	5	1
18	5	3	5	2
19	5	3	5	3
20	5	3	5	4

Loc	Moy
1	3
2	3
3	3
4	3

Valeurs inférées par l'environnement - cas 1



# ON DISTRIBUE LA POPULATION AUTREMENT...



La science pour la santé  
From science to health

Id	V1	V2	Loc
1	1	1	1
2	1	1	2
3	1	1	3
4	1	1	4
5	2	2	1

Id	V1	V2	Loc
1	1	1	1
2	1	1	1
3	1	1	4
4	1	1	4
5	2	2	2

Id	V1	V2	Loc
1	1	1	1
2	1	1	1
3	1	1	2
4	1	1	3
5	2	2	1

Loc	Moy
1	3
2	3
3	3
4	3

Loc	Moy
1	3,17
2	2,75
3	2,75
4	3,60

Loc	Moy
1	1,33
2	2,40
3	3,29
4	4,20

15	4	4	3
16	4	4	4
17	5	5	1
18	5	5	2
19	5	5	3
20	5	5	4

15	4	4	3
16	4	4	4
17	5	5	1
18	5	5	1
19	5	5	4
20	5	5	4

15	4	4	4
16	4	4	4
17	5	5	3
18	5	5	3
19	5	5	4
20	5	5	4



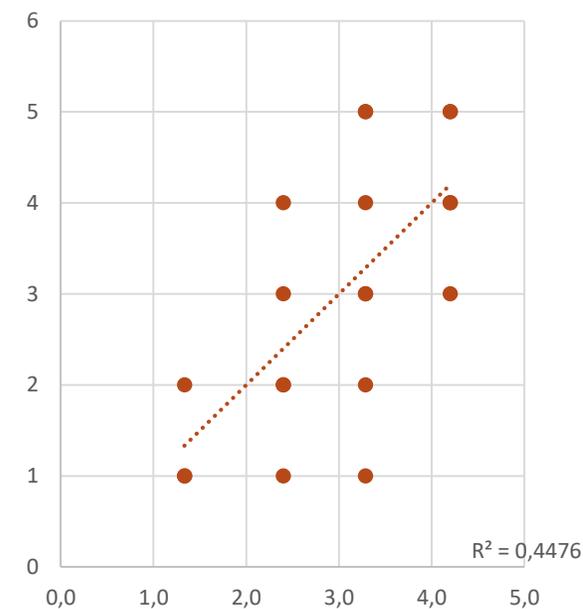
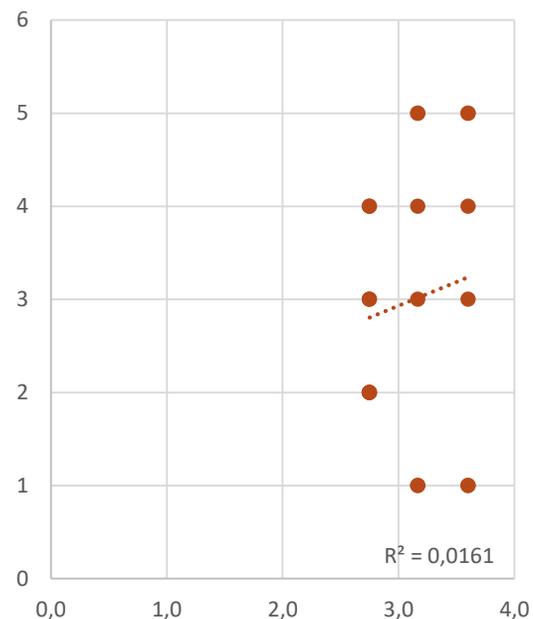
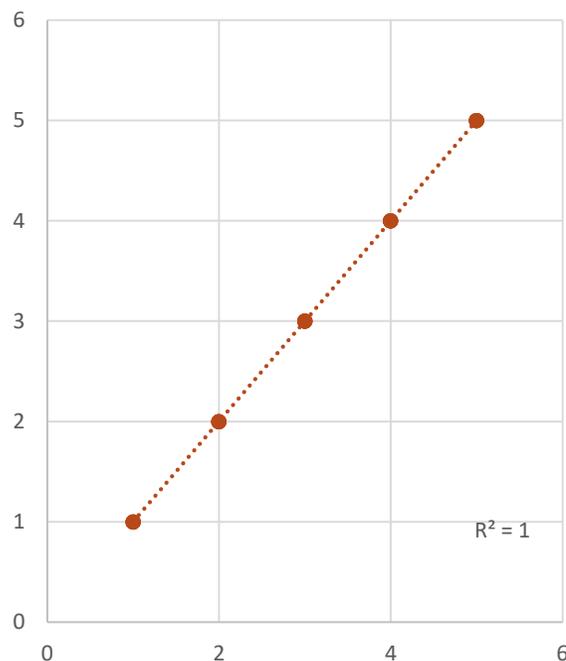
# ON PERD LA CORRÉLATION !



Valeurs inférées par  
l'environnement - cas 2

Valeurs inférées par  
l'environnement - cas 3

Valeurs réelles



slido

Rejoindre sur  
**slido.com**  
**#Ethique**



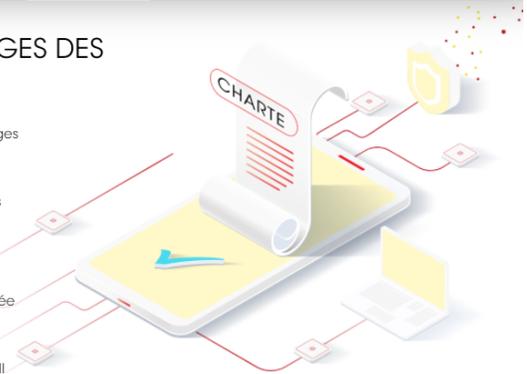
# L'éthique, un pilier d'Occitanie Data

- Commission éthique animée par Emmanuelle Rial-Sebbag
- Une cheffe de projet Ethique de la donnée
- Des réalisations :
  - Charte éthique
  - KanopyMed
  - Projet PEP

## LA CHARTE ÉTHIQUE DES USAGES DES DONNÉES

Afin de proposer un cadre de confiance pour les échanges et les partages de données entre structures publiques, privées et académiques, Occitanie Data a entrepris dès l'été 2019 d'élaborer une Charte éthique des usages des données.

Une première version de cette Charte (à télécharger ci-dessous) est née d'une collaboration fructueuse entre Occitanie Data et l'équipe de recherche BIOETHICS, dirigée par Emmanuelle Rial-Sebbag (une équipe de recherche spécialisée dans les questions éthiques liées aux innovations en santé et rattachée à l'Université Toulouse III Paul Sabatier ainsi qu'à l'INSERM).



Trouvez l'avenir qui vous correspond

JE ME LANCE ! >



Commencer...



Découvrir

Quelles sont mes compétences ?



Explorer



Optimiser

Quel parcours suivre ?

DONNÉES PERSONNELLES

ETHIQUE

Le service « Parcours Emploi Personnalisé », proposé par la Région Occitanie dans le cadre du pacte vert, s'inscrit dans le cadre de confiance porté par Occitanie Data « Data et IA en pleine confiance » et sa charte éthique de la donnée. Ce cadre apporte l'assurance que les données qui serviront à des traitements algorithmiques le seront dans un cadre juridique solide, avec une éthique garantie par un processus de contrôle indépendant, pour des usages respectueux de principes convenus par tous, notamment en matière environnementale, et à des fins d'intérêt général.

Ce cadre repose ainsi sur des principes éthiques, exposés ci-dessous. La Région Occitanie confronte en continue le service « Parcours Emploi Personnalisé » à ces principes. Ce faisant la Région Occitanie garantit à tous, et en premier lieu aux citoyens, que si la collectivité recourt à des outils performants de gestion des données pour améliorer le service rendu aux usagers, elle le fait dans une approche garantissant à terme des choix durables, éthiques, équitables et souverains.

- Concept note : **Advancing data justice research and practice**

**High-Level Goal:** This project aims to fill a gap in Data Justice research and practice that provides a frame to help practitioners and users to move beyond understanding data governance narrowly, as a compliance matter of individualised privacy or ethical design, to include **considerations of equity and justice in terms of access to, and visibility and representation in, data** used in the development of artificial intelligence / machine learning (AI/ML) systems. The project aims to provide (a) a assessment of the current state of research in this area and the identification of gaps and (b) a preliminary guide for the three types of stakeholders (i.e., policymakers, developers, and communities affected by AI/ML systems) to consider in their practice and use of AI/ML systems and specifically in consideration of realising the 2030 Sustainable Development Goals.

- Concept note : **Enabling data sharing for social benefit through data institutions**

The Data Trusts project will support the creation of real-world data trusts that enable data sharing for social benefit. In so doing, it will provide new institutions that empower individuals and communities to enact their data rights, ensuring that data sharing activities reflect the diverse interests of all in society.

In pursuit of this goal, the project will help build the foundations for pilot studies and generate toolkits that demonstrate the value of data trusts and how they can operate in practice, creating the conditions in which others in society are able to set up a data trust to meet their needs.

Merci pour votre attention