

Anonymisation :

# Êtes-vous à l'abri de sanctions ?



Louis Philippe  
**SONDECK, PhD**

CEO & Founder  
Docteur en anonymisation des  
données

## Contacts

[louis.sondeck@cleveridentity.fr](mailto:louis.sondeck@cleveridentity.fr)

[linkedin.com/in/cleveridentity](https://www.linkedin.com/in/cleveridentity)

# CLEVER IDENTITY

Assess Your Re-identification Risks



Pure player de l'anonymisation /  
Pseudonymisation  
de données

- Editeur de solutions
- Conseil et formation (DPO et SSI)

# PROGRAMME

1. Les enjeux de l'anonymisation
2. Définitions
3. Solutions concurrentes et limites
4. La proposition de valeur Clever Identity
5. Questions et réponses



# Les enjeux de l'anonymisation



Les enjeux de l'anonymisation

# L'HISTOIRE DE BERNARD



Bernard

Emprunt Immobilier



Banque

Crédit de 20k

Crédit de 25k

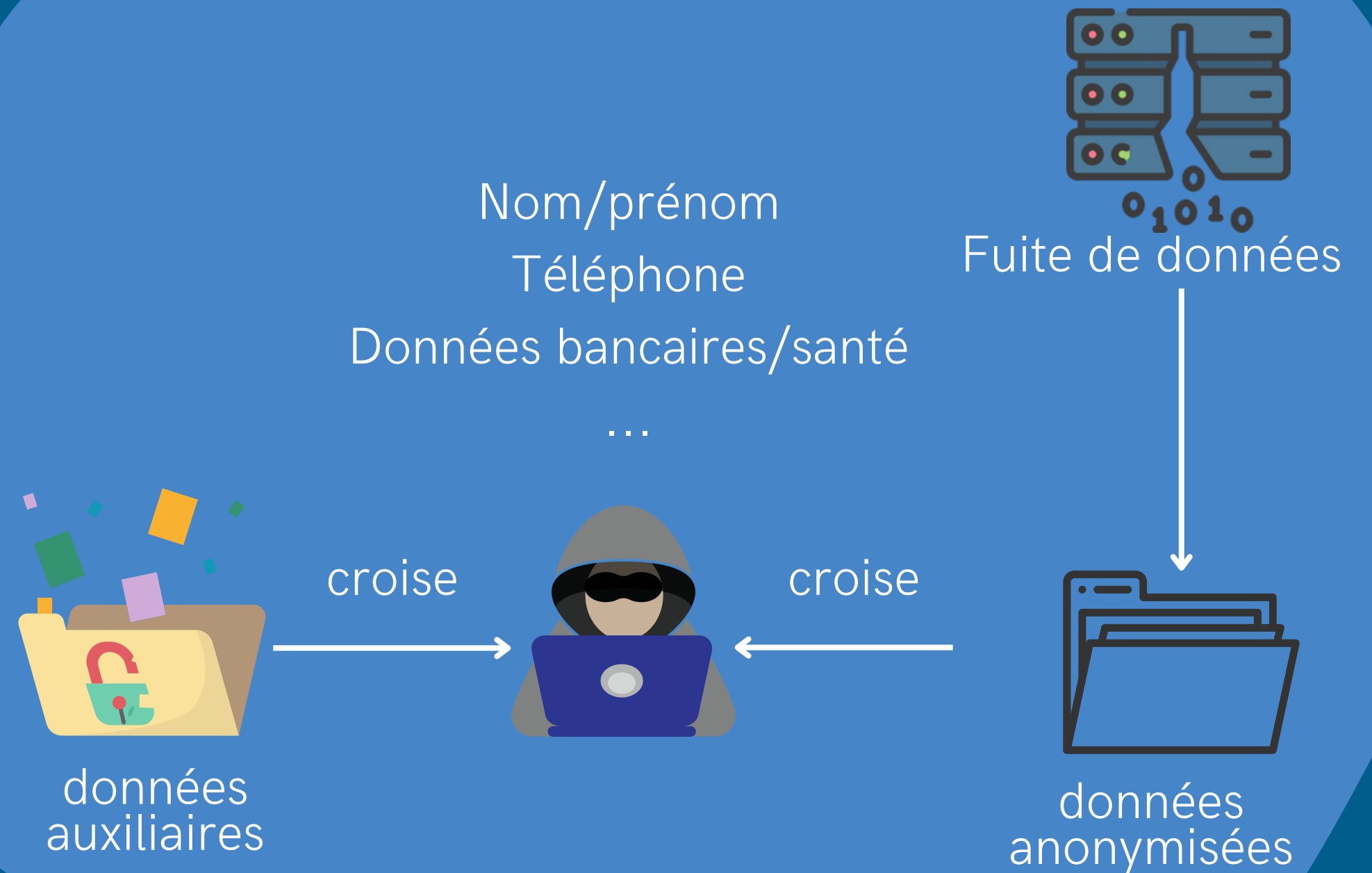
Chèque impayé de 32k

Fich à la banque de France



Les enjeux de l'anonymisation

# L'HISTOIRE DE BERNARD



Les enjeux de l'anonymisation

# L'HISTOIRE DE BERNARD

Données  
auxiliaires

Noms	Adresse mail	Age	Sexe	Code Postal	Nombre de clics
Jack Sparrow	jacks@yahoo.fr	26	M	75003	15
Kevin Mitnick	mariew@hotmail.com	36	F	75012	22
Bernard Olivier	berno@yahoo.com	27	M	75008	30

Données  
"anonymisées"

Matricule	Adresse mail	Noms	Téléphone	Age	Sexe	Code Postal	Maladie	Revenus
12AD345B	**@yahoo.fr	*****	*****	26	M	75003	VIH	75K
456A89DB	**@hotmail.com	*****	*****	36	F	75012	Cancer	25K
99CDF234	**@yahoo.com	*****	*****	27	M	75008	VIH	100K
F567A235	**@gmail.com	*****	*****	32	F	75013	Diabètes	30K



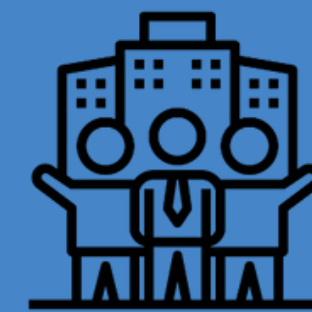
Les enjeux de l'anonymisation

# LES IMPACTS



Bernard

Psychologique  
Financier  
Réputation  
Règlementaire



L'entreprise

Image  
Financier  
Organisationnel  
Juridique  
Règlementaire





Les enjeux de l'anonymisation

# SITUATIONS À RISQUE

**Conservation des données au-delà de la durée légale**

Non respect des durées de conservation (RGPD Art. 5 et Loi 78-17 Art. 4)

**Open Data**

Respect du Code des relations entre le public l'administration (article L. 312-1-2)

**Recherche en santé (EDS...)**

Echange de données entre espaces de travail ou partage avec des partenaires externes (Ref. EDS)

**Data Lab**

Absence du droit d'en connaître des Data Scientist

**Environnements hors production (Dev, Recette...)**

- Environnements peu sécurisés en général (RGPD Art. 32)
- Absence du droit d'en connaître des développeurs



Les enjeux de l'anonymisation

# Sur le plan réglementaire

2019  
600 K€

2020  
2.3 M€

2021  
1.8 M€

Sanctions de la CNIL liées à la  
sécurité des données

2022  
1.8 M€

- Délibération n° 2015-255 (Anonymisation non conforme - SHA256)
- Délibération n° SAN-2021-008 (Anonymisation non conforme)
- Délibération n° 2016-047 (Anonymisation conforme)
- Délibération n° 2017-337 (Anonymisation conforme)

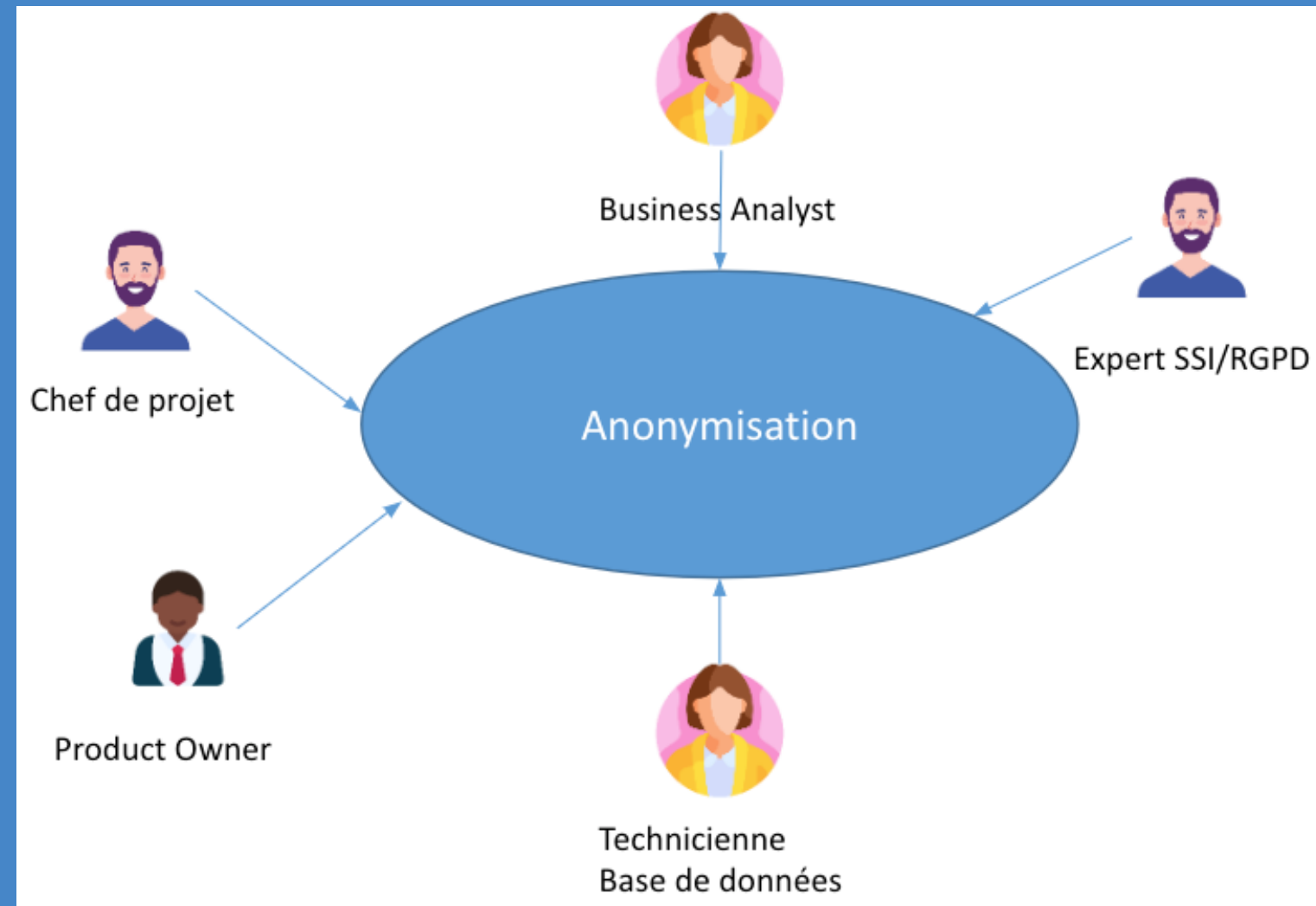


Les enjeux de l'anonymisation

# Sur le plan organisationnel

## Perte de temps

- Ateliers chronophages
- Processus non industrialisé (fichiers excel)
- Difficultés à faire évoluer les outils (nouvelles méthodes, adaptation des méthodes au besoin métier...)
- Faible disponibilité des équipes techniques



Les enjeux de l'anonymisation

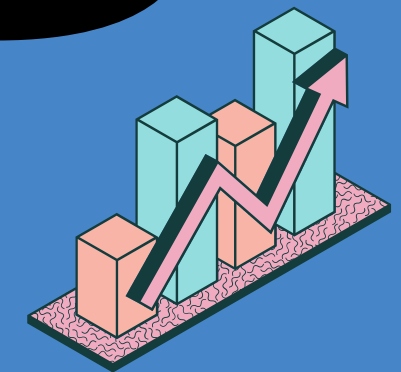
# Sur le plan technique



Protection de la  
vie privée



Equilibre



Valeur  
ajoutée pour  
l'entreprise



# Définitions



## Définitions

# ANONYMISATION VS PSEUDONYMISATION

Processus	Pseudonymisation	Anonymisation
<b>Statut des données</b>	Personnelles (restent indirectement identifiantes et donc soumises au RGPD et à la loi Informatique et Libertés)	Anonymes
<b>Réutilisation des données</b>	Sous conditions	Sans restriction
<b>Utilité des données</b>	Préservée car pas d'altération du niveau de détail des données	Plus ou moins altérée en fonction des objectifs poursuivis et des méthodes appliquées
<b>Méthodes à mettre en œuvre</b>	Compteur, générateur de nombres aléatoires, fonction de hachage, chiffrement à clé secrète, etc.	Dépend des objectifs poursuivis : confidentialité différentielle, randomisation, k-anonymat, l-diversité, t-proximité, etc.
<b>Complexité de la mise en œuvre</b>	Simple à moyenne	Dépend des objectifs poursuivis : simple dans certains cas comme l'agrégation ou le comptage et complexe dans d'autres

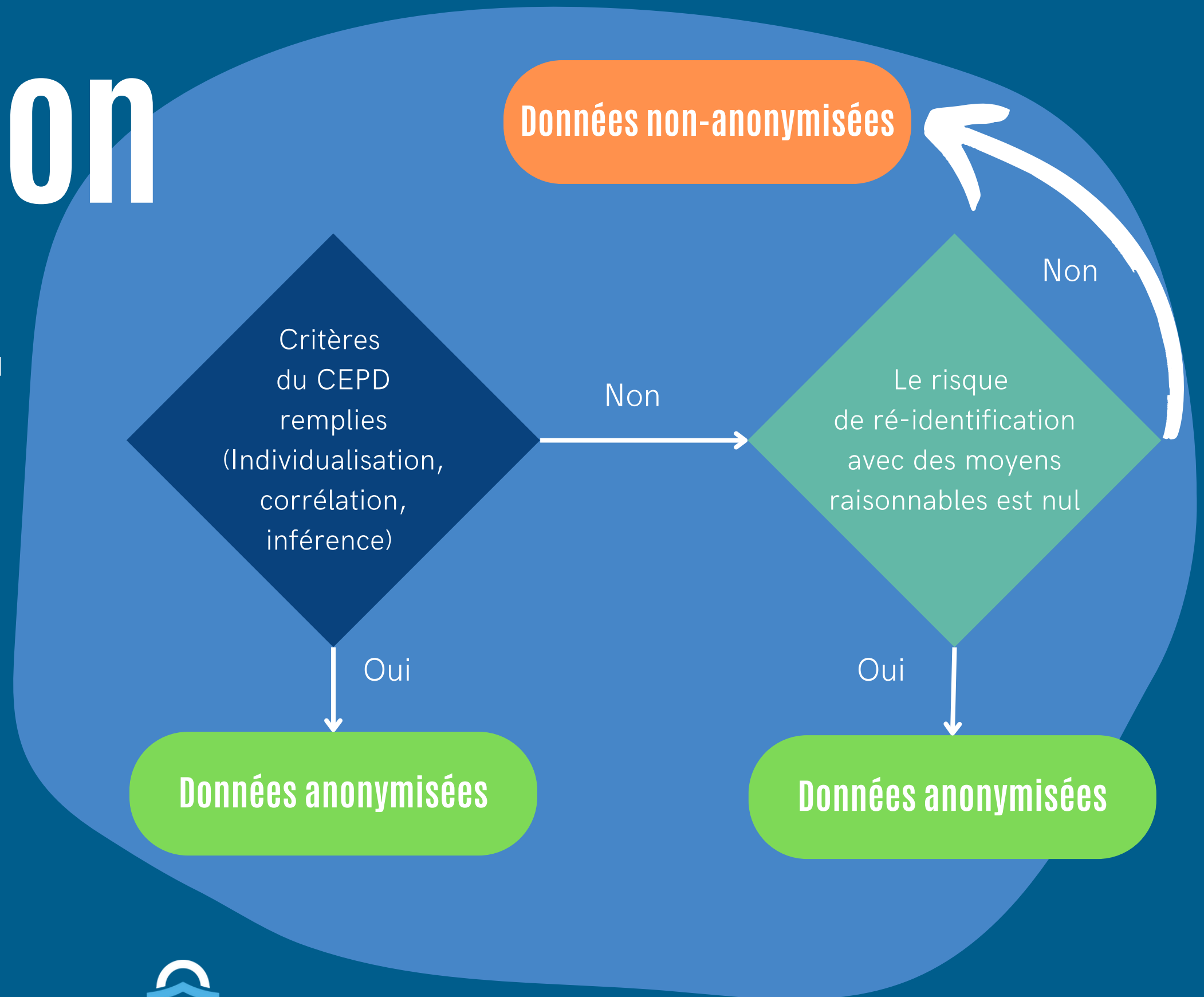


## Définitions

# Anonymisation selon la CNIL

À défaut de remplir parfaitement ces trois critères, il doit être démontré, via une évaluation approfondie des risques d'identification, que le risque de ré-identification avec des moyens raisonnables est nul.

LA CNIL



## Définitions

# L'INDIVIDUALISATION



C'est la possibilité d'isoler un individu dans le jeu de données

Avis du G29

Données "anonymisées"

Matricule	Noms	Adresse mail	Age	Sexe	Code Postal	Maladie
12AD345B	*****	**@yahoo.fr	30	M	75***	VIH
456A89DB	*****	**@hotmail.com	40	F	75***	Cancer
99CDF234	*****	**@yahoo.com	22	M	75***	VIH
F567A235	*****	**@gmail.com	33	F	75***	Diabètes





Définitions

# L'INFÉRENCE



C'est la possibilité de déduire, de façon quasi-certaine, de nouvelles informations sur un individu

Avis du G29

Données "anonymisées"

Matricule	Noms	Adresse mail	Age	Sexe	Code Postal	Maladie
12AD345B	*****	**@yahoo.fr	30	M	75***	VIH
456A89DB	*****	**@hotmail.com	40	F	75***	Cancer
99CDF234	*****	**@yahoo.com	22	M	75***	VIH
F567A235	*****	**@gmail.com	33	F	75***	Diabète



Définitions

# LA CORRÉLATION



C'est la possibilité de relier entre eux des ensembles de données distincts concernant un même individu

Avis du G29

Données auxiliaires

Noms	Adresse mail	Sexe	Code Postal
J.Sparrow	js@yahoo.fr	M	75003
M.Mbog	mb@hotmail.com	F	75008
L.Fary	lf@yahoo.com	F	75008

Données "anonymisées"

Matricule	Noms	Adresse mail	Age	Sexe	Code Postal	Maladie
12AD345B	*****	**@yahoo.fr	30	M	75***	VIH
456A89DB	*****	**@hotmail.com	40	F	75***	Cancer
99CDF234	*****	**@yahoo.fr	22	M	75***	Cancer
F567A235	*****	**@gmail.com	33	F	75***	Diabètes



# Solutions concurrentes et limites



Solutions concurrentes et limites

# 2 GRANDES FAMILLES DE SOLUTIONS

## Data Masking (>95% des cas)

### Avantage

- Réutilisation facile des données

### Inconvénients

- Pseudonymisation
- Aucune évaluation selon les 3 critères

## Données synthétiques (<5% des cas)

### Avantage

- Données artificielles

### Inconvénients

- Coût élevé
- Chronophage (des délais d'un mois pour une table de données...)
- Qualité des données non garantie



# Proposition de valeur Clever Identity



# LE MEILLEUR DES 2 MONDES



## CleverConf

- Evaluation de la sensibilité des données (sensibles, identifiants, quasi-identifiant...)
- Calibrage de l'anonymisation (équilibre entre vie privée et qualité des données)
- 



## CleverAnon

- Méthodes recommandées (k-anonymisation, l-diversité, ajout de bruit...)
- Deux modules : data et log
- Automatisation



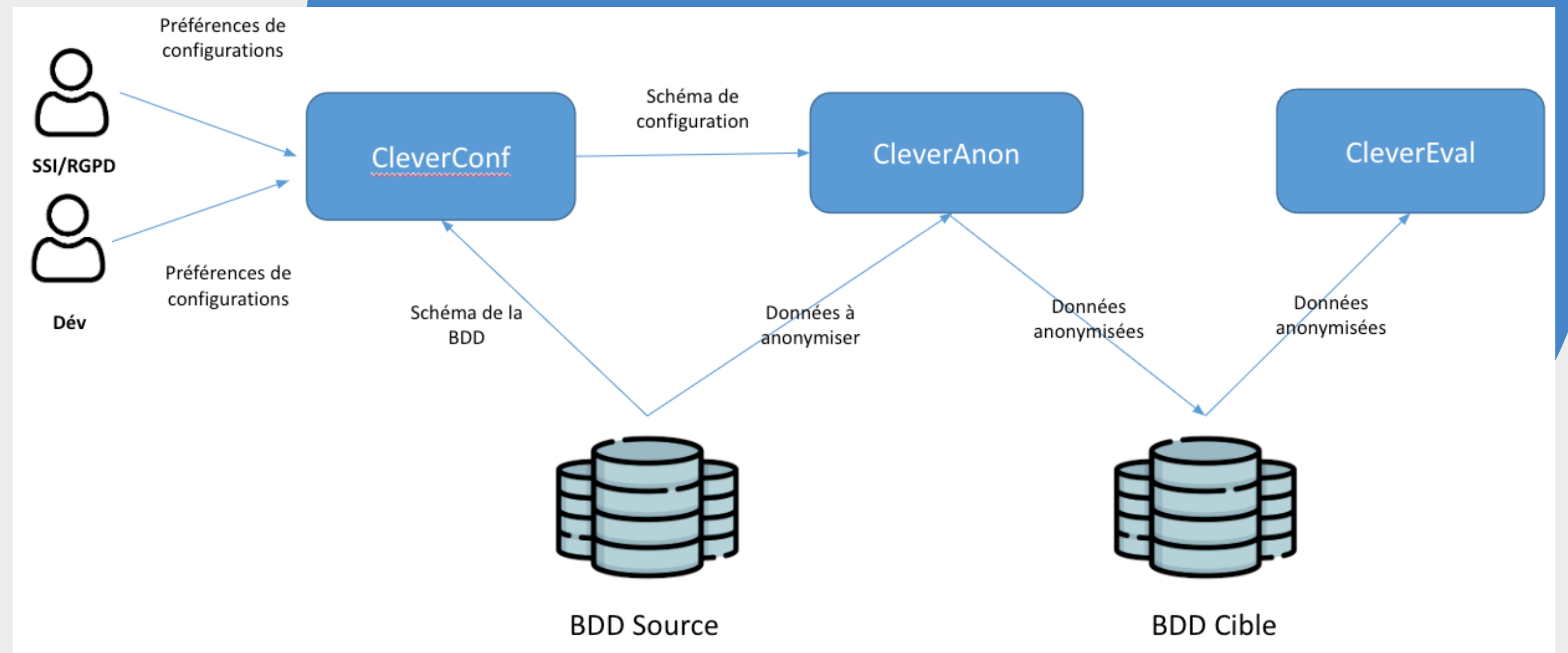
## CleverEval

- Evaluation des 3 critères (scan de vulnérabilités)
- Evaluation contextuelle
- Remédiation automatisée
- Dossier de preuves

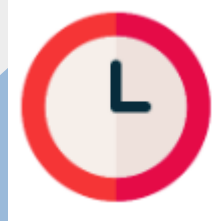


Proposition de valeur de Clever Identity

# Architecture Outils Clever

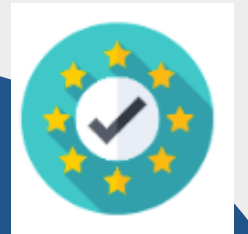


# Avantages des outils Clever



## Gain de temps

- Automatisation et industrialisation des processus
- Evolution aisée
- Réutilisation des modèles
- Peu ou pas de réunions projet
- Peu ou pas d'intervention technique



## Conformité et Auditabilité

- Evaluation selon les critères du G29
- Méthodes conformes (généralisation, randomisation)
- Suivi aisé des modifications
- Dossier de preuves





Proposition de valeur de Clever Identity

# Partenariat et contacts

Innovation



Contacts

Contactez-nous :  
[info@cleveridentity.fr](mailto:info@cleveridentity.fr)  
06 51 30 25 97  
[linkedin.com/in/cleveridentity](https://www.linkedin.com/in/cleveridentity)



[Cleveridentity.fr](https://www.cleveridentity.fr)



# QUESTIONS & REPOSSESSES