PostgreSQL et le principe de Privacy By Design

Table des matières

Bonjour

- Damien Clochard
- DBA PostgreSQL & Co-foundateur de Dalibo
- Président de l'association PostgreSQLFr
- Je ne suis pas juriste !

Mon chemin

Menu

- RGPD : 3 an plus tard…
- Pourquoi c'est difficile ?
- Protéger les données dès la conception
- PostgreSQL Anonymizer

RGPD

- Droits Individuels
- Principes
- Impact

RGPD : les droits Individuels

- droit à l'information (Art. 13 et Art. 14)
- droit d'accès (Art. 15)
- droit de rectification (Art. 16)
- droit à la portabilité (Art 20)
- droit d'opposition (Art. 21)
- **droit à l'oubli** (Art. 17)
- **droit à la limitation du traitement** (Art. 18)
- droit de décision automatisée (Art. 22)

(sources: Individual Rights[1])

RGPD: Principes & Concepts

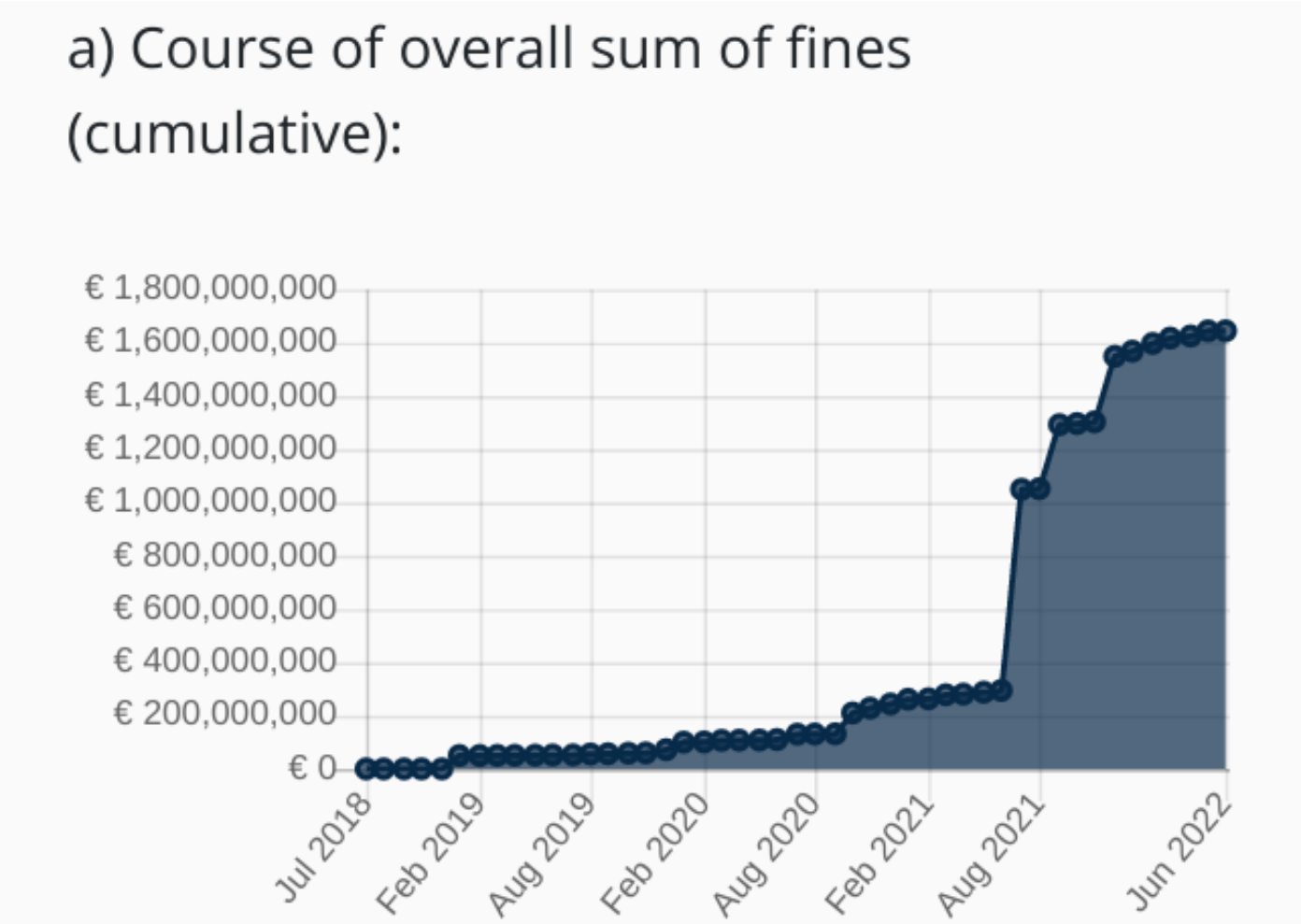
- Licéité, loyauté, transparence
- Sécurité
- Minimisation des données
- **Privacy By Design**
- **Data Protection By Design**
- Limitation du stockage
- Précision
- Limitation des finalités

(source: GDPR Principles[2])

---

[1] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/
[2] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/

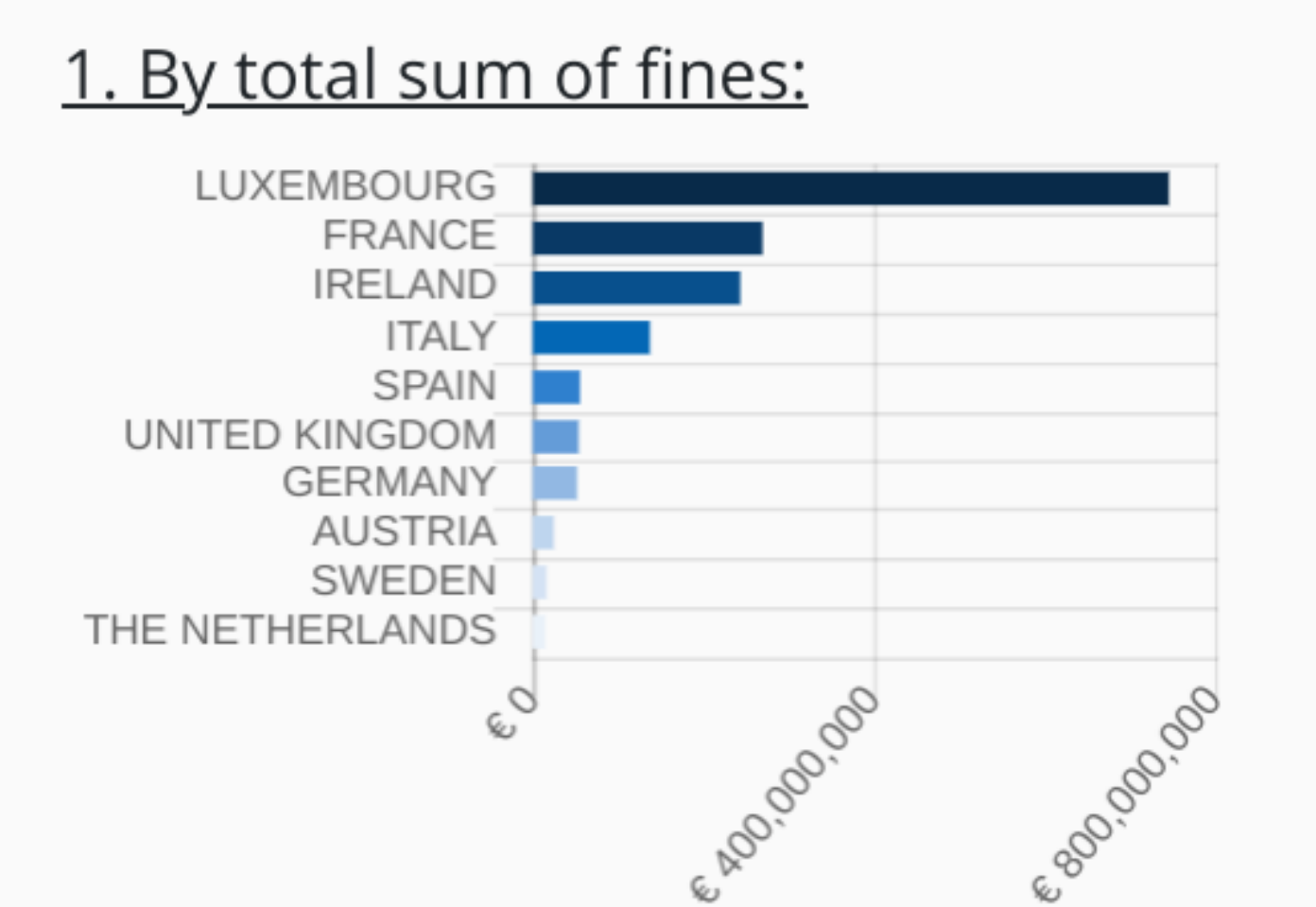
Les montants explosent



(source: GDPR Enforcement Tracker[3])

[3] http://www.enforcementtracker.com/

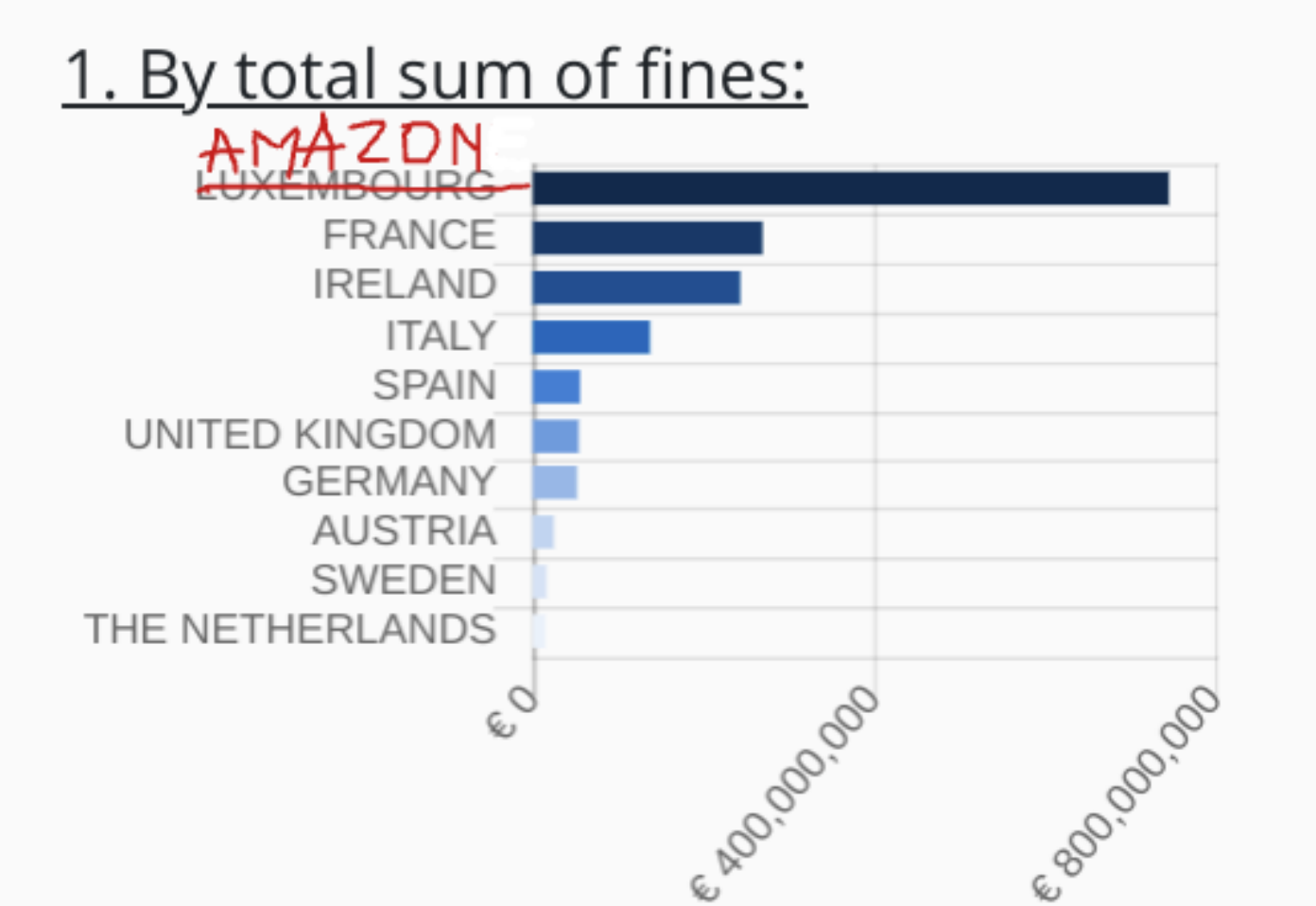
La France est le mauvais élève



(source: GDPR Enforcement Tracker[4])

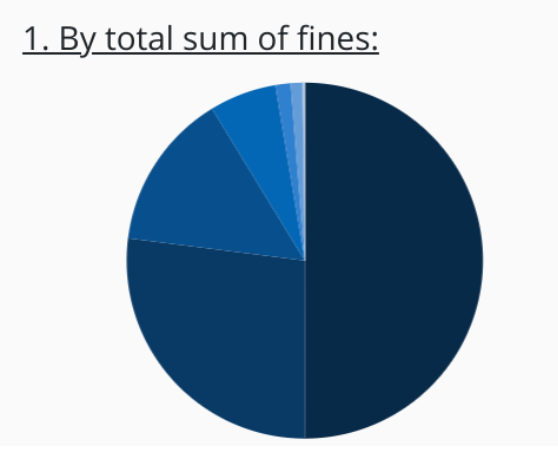
[4] http://www.enforcementtracker.com/

La France est le mauvais élève



(source: GDPR Enforcement Tracker[5])

Les principes RGPD sont aux coeurs des sanctions



	Violation	Sum of Fines
	Non-compliance with general data processing principles	€ 824,981,574 (250 fines)
	Insufficient legal basis for data processing	€ 446,465,031 (393 fines)
	Insufficient fulfilment of information obligations	€ 235,796,695 (96 fines)

[5] http://www.enforcementtracker.com/

(source: GDPR Enforcement Tracker[6])

[6] http://www.enforcementtracker.com/

POURQUOI C'EST DIFFICILE ?

En général:

- L'anonymisation est faite en bout de chaine
- La surface d'attaque est trop grande
- Les développeurs/éditeurs ne sont pas impliqués
- Les outils d'anonymisation sont externes

Le RGPD a identifié le problème

Article 25[7]

« […] le responsable du traitement met en œuvre, **tant au moment de la détermination des moyens du traitement** qu'au moment du traitement lui-même, des mesures techniques et organisationnelles appropriées […] »

---

7 Bonnes Pratiques d'anonymisation

- Embarquer les règles d'anonymisation
- Privacy By Default
- Qualifier les roles
- Anonymiser dans la base
- Suivre le cycle de vie des données
- Echantillonner
- Evaluer

---

[7] https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre4#Article25

CONCRÈTEMENT ?

[8]

PostgreSQL Anonymizer

- Extension open-source pour PostgreSQL
- Fonctionne avec toutes les versions
- (… mais pas sur Amazon RDS)
- Moteur de masquage + boite à outil
- version 1.0 sortie en mai
- https://labs.dalibo.com/postgresql_anonymizer

Exemple

```
CREATE TABLE customer (
    id SERIAL PRIMARY KEY,
    firstname TEXT,
    lastname TEXT,
    phone TEXT,
    birth DATE,
);
```

[8] https://postgresql-anonymizer.readthedocs.io/en/stable/

Embarquer les règles d'anonymisation

```
SECURITY LABEL FOR anon ON COLUMN customer.lastname
IS 'MASKED WITH FUNCTION anon.fake_last_name()';
```

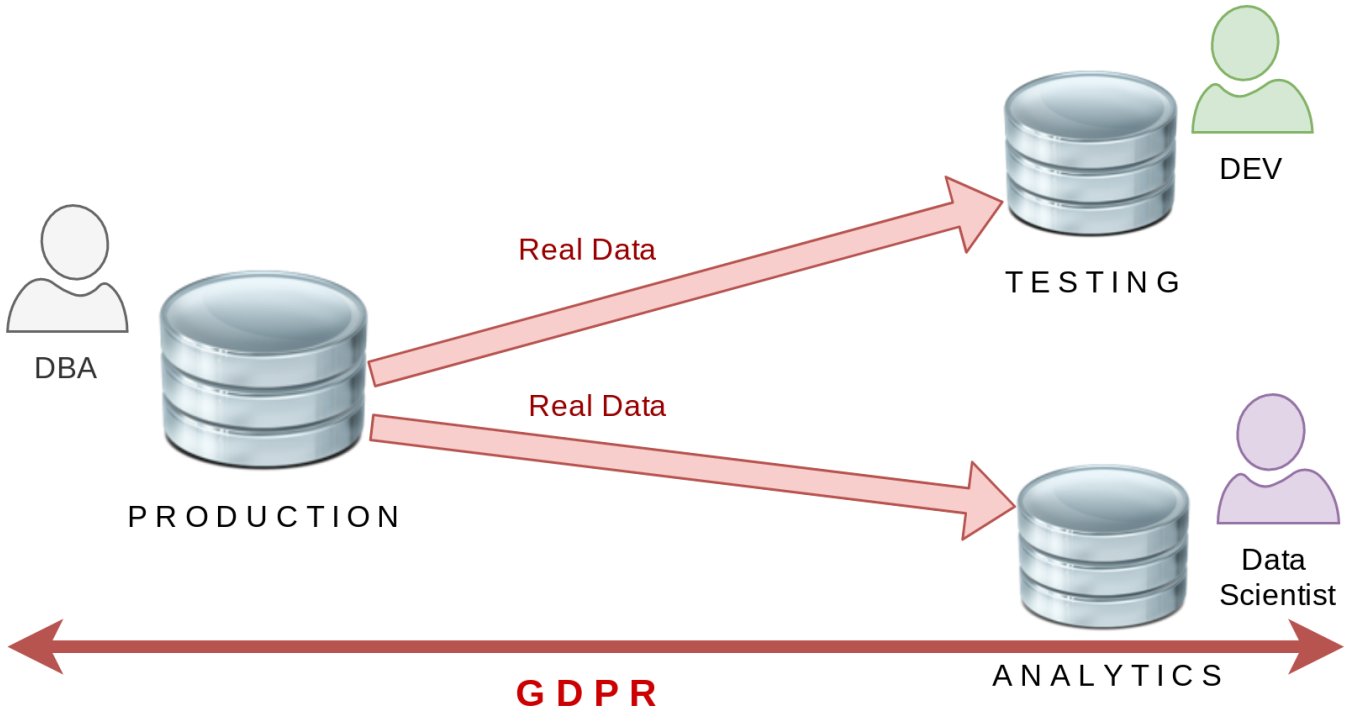
Privacy By Design

```
SECURITY LABEL FOR anon ON COLUMN customer.phone
IS 'MASKED WITH VALUE $$CONFIDENTIAL$$';
```

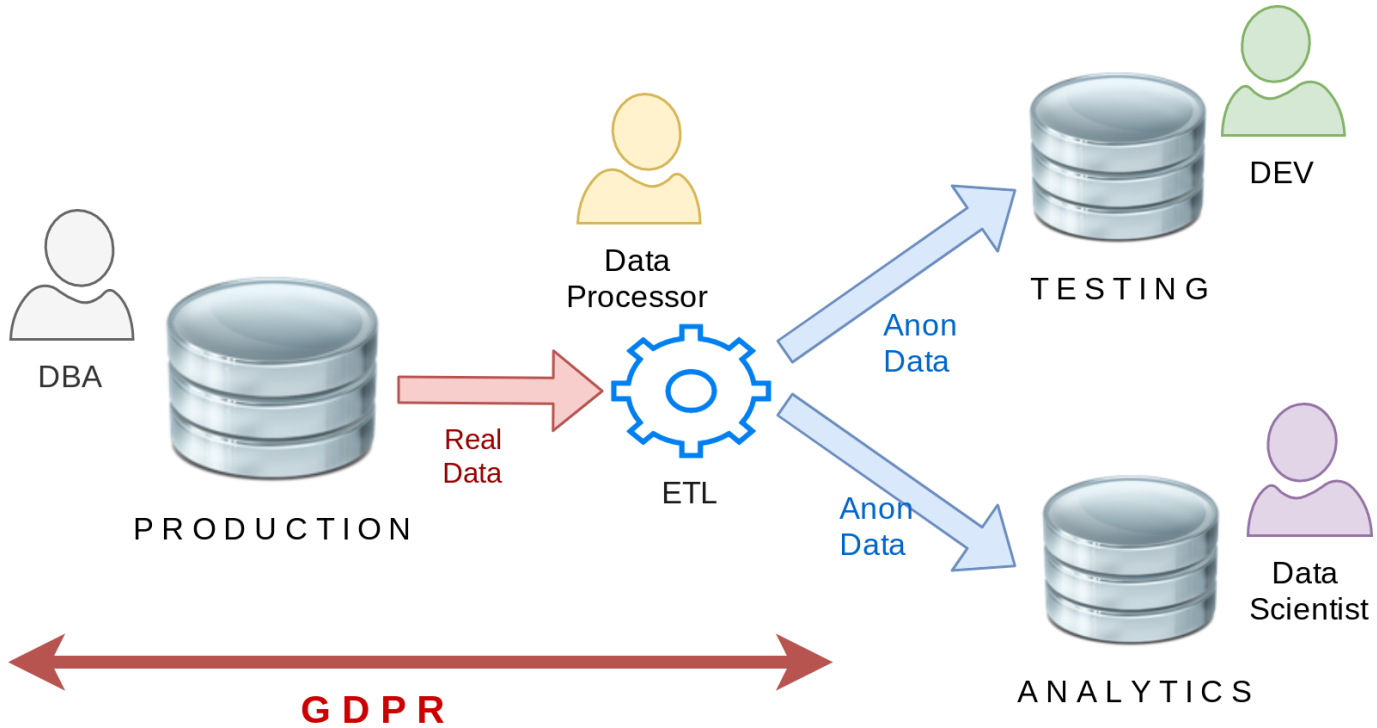
Qualifier les roles

```
SECURITY LABEL FOR anon ON COLUMN data_scientist
IS 'MASKED'
```

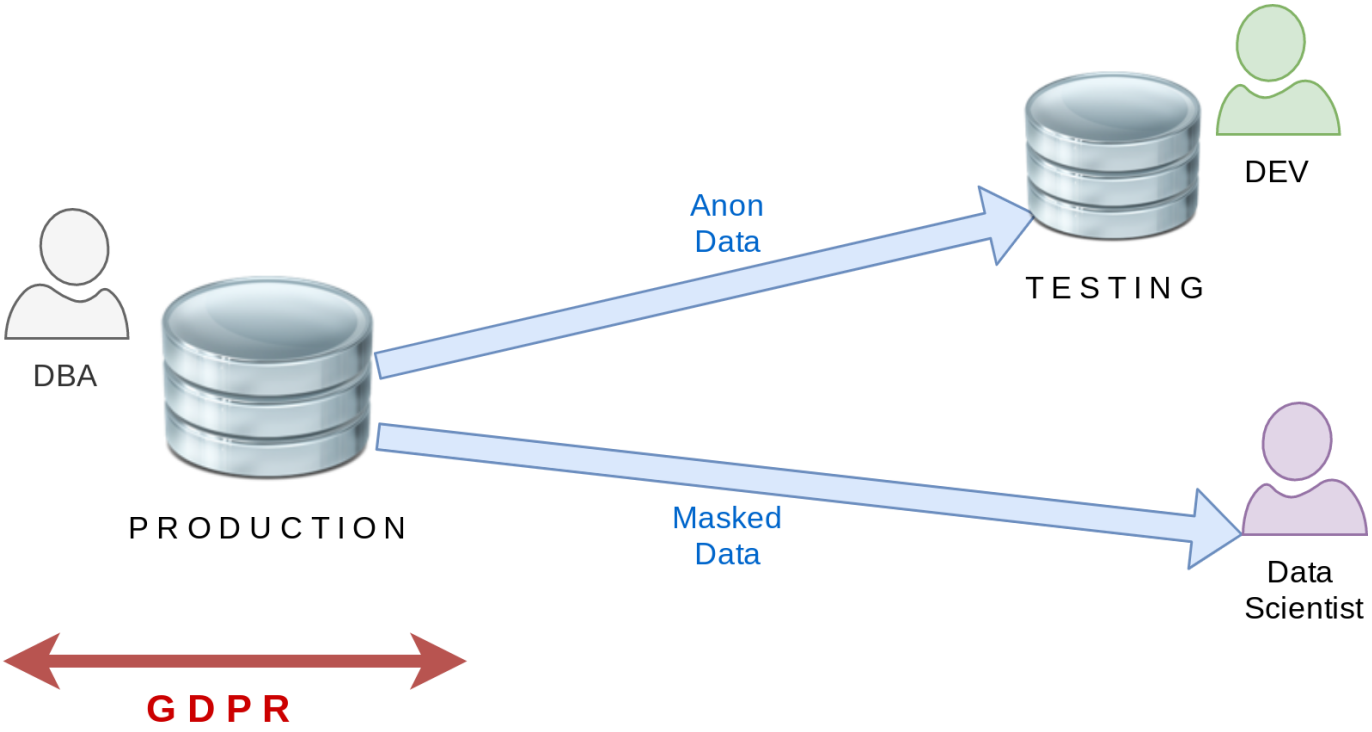
Anonymiser dans la base



Anonymiser dans la base



Anonymiser dans la base



Suivre le cycle de vie des données

```
ALTER TABLE customer ADD COLUMN postcode TEXT;

SECURITY LABEL FOR anon ON COLUMN customer.postcode
IS 'MASKED WITH VALUE NULL';
```

Echantillonner

```
SECURITY LABEL FOR anon ON TABLE customer
IS 'TABLESAMPLE BERNOULLI 33';
```

PS: cette fonction est en cours de développement :-)

Evaluer

```
SECURITY LABEL FOR anon ON COLUMN customer.birth
IS 'INDIRECT IDENTIFIER';
SECURITY LABEL FOR anon ON COLUMN customer.postcode
IS 'INDIRECT IDENTIFIER';
```

```
SELECT anon.k_anonymity('customer')
   k_anonymity
-----------------
 3
```

EN RÉSUMÉ

- Les sanctions du RGPD sont bien réelles
- Les fuites de données sont le plus gros risque
- Reduire la surface d'attaque
- Anonymiser dès que possible
- Anonymiser dans la base de données

Bataille pour la vie privée

- Les développeurs doivent écrire les règles de masquage
- C'est difficile mais PostgreSQL est un bon point de départ
- La protection des données privées est un travail d'équipe
- Les éditeurs doivent livrer les règles de base d'anonymisation

---

Aller plus loin

- Worshop[9]
- Video[10]

Comment Contribuer ?

- Feedback et bugs !
- Témoignages
- Rejoindre le projet sur :

https://gitlab.com/dalibo/postgresql_anonymizer

---

[9] https://dalibo.gitlab.io/postgresql_anonymizer/how-to.handout.pdf
[10] https://www.youtube.com/watch?v=siUz_W93A9U

Merci !

DGFIP

Biomerieux

Mes collègues

A bientôt !

- Contact : damien.clochard@dalibo.com[11]
- Follow : @daamien[12]
- Nos autres Projets : Dalibo Labs[13]

[11]

[12] https://twitter.com/daamien
[13] https://labs.dalibo.com