PostgreSQL et le principe de Privacy By Design

DALIBO
L'expertise PostgreSQL

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)

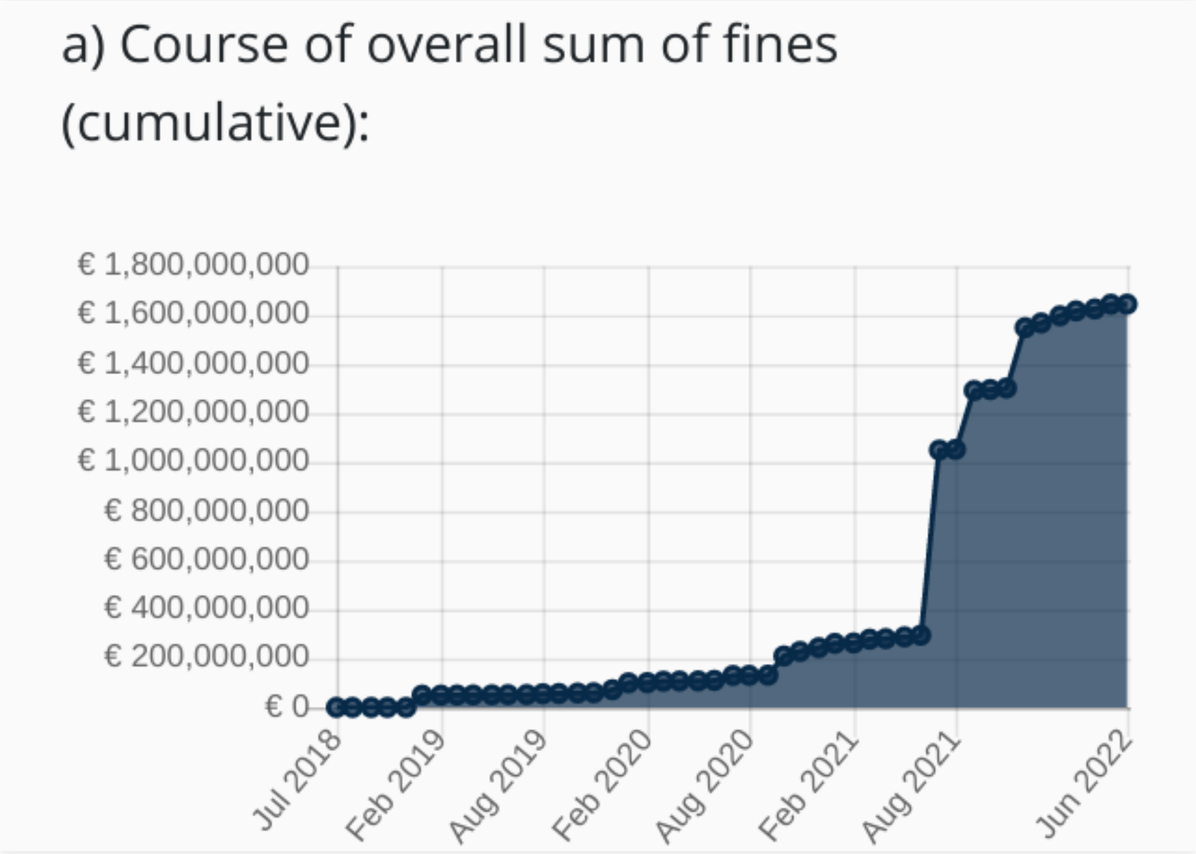
RGPD: PRINCIPES & CONCEPTS

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

(source: GDPR Principles)

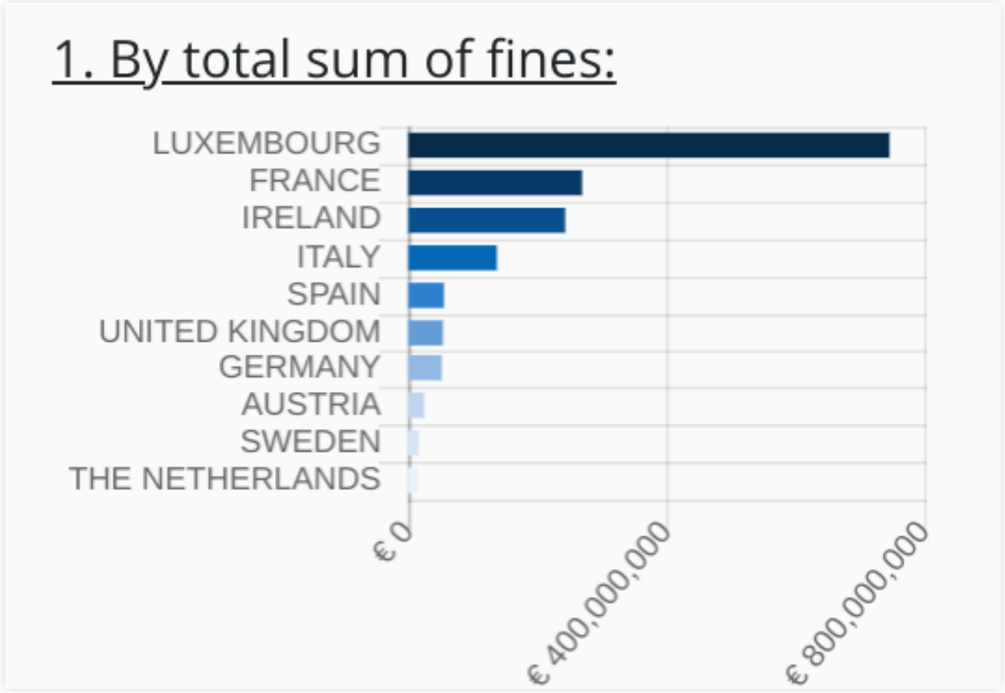

GDPR SANCTIONS ARE COMING

LES MONTANTS EXPLOSENT



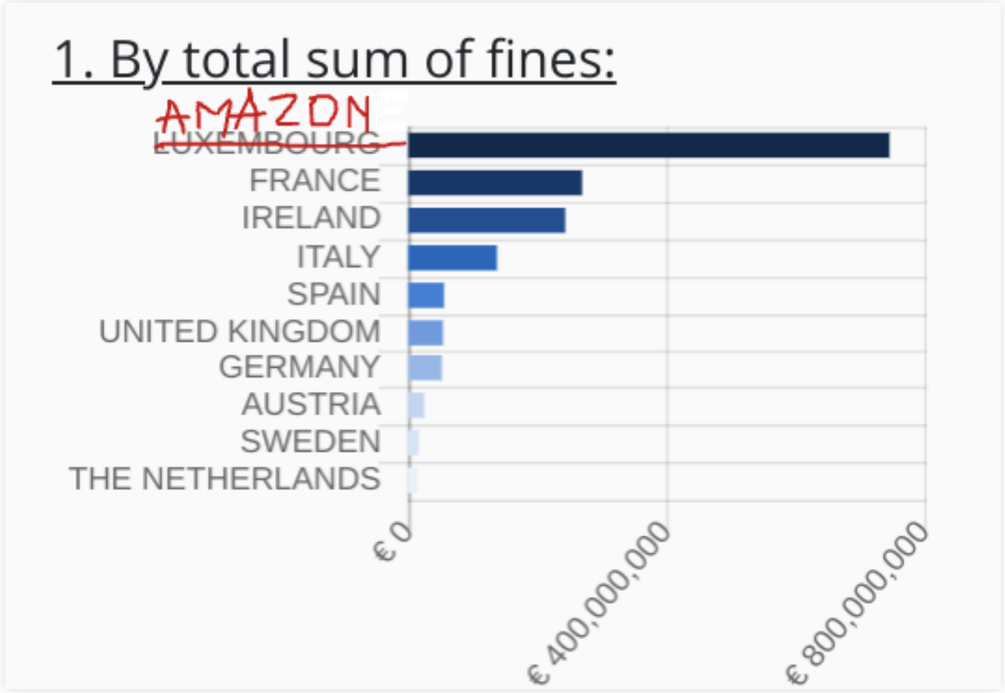
(source: GDPR Enforcement Tracker)

LA FRANCE EST LE MAUVAIS ÉLÈVE



(source: GDPR Enforcement Tracker)

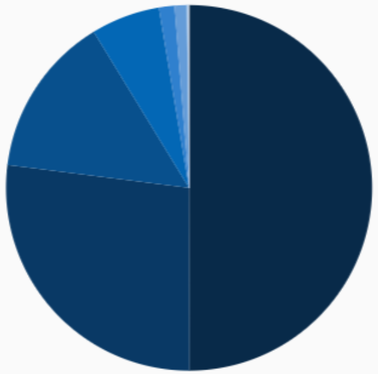
LA FRANCE EST LE MAUVAIS ÉLÈVE



(source: GDPR Enforcement Tracker)

LES PRINCIPES RGPD SONT AUX COEURS DES SANCTIONS

1. By total sum of fines:



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)

(source: GDPR Enforcement Tracker)

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

« […] 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

Concrètement ?


PostgreSQL
Anonymizer

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,
);
```

EMBARQUER LES RÈGLES D'ANONYMISATION

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

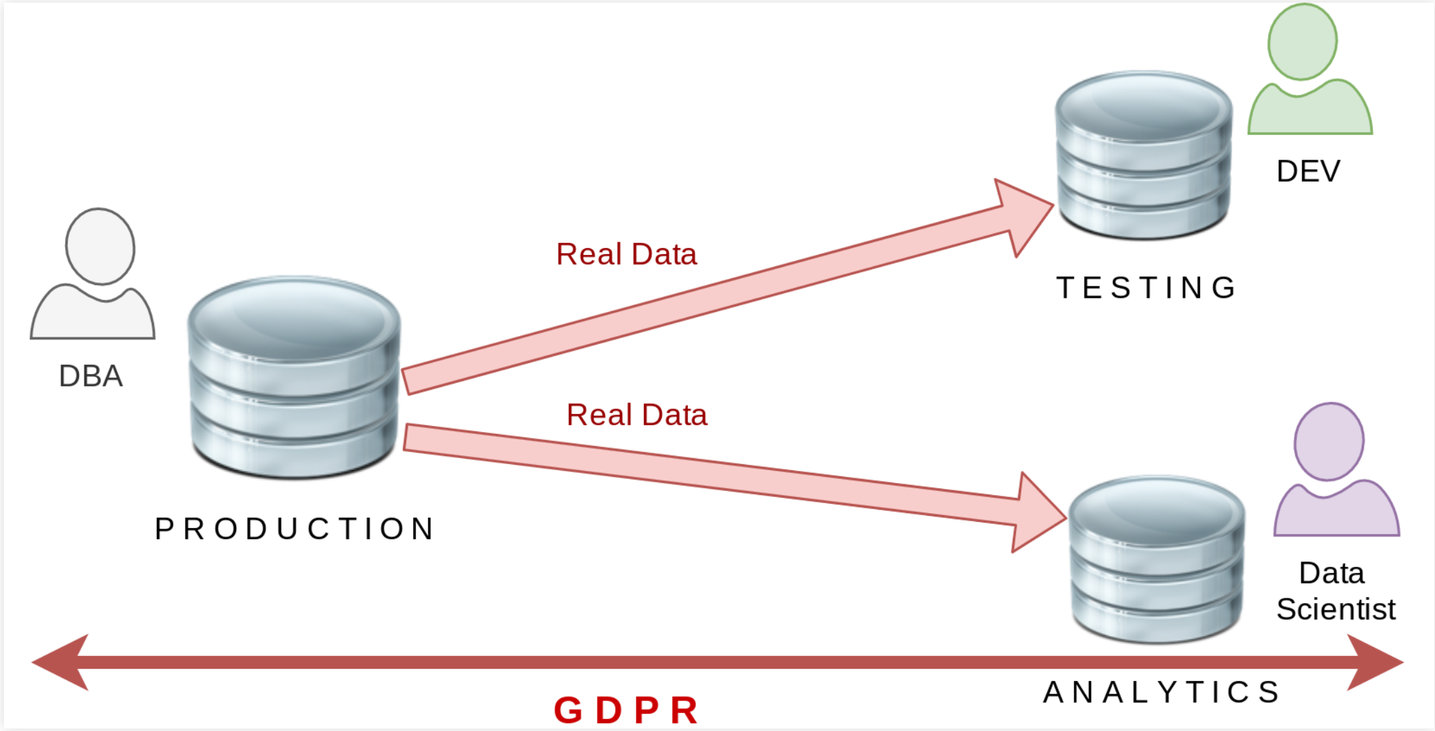
PRIVACY BY DESIGN

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

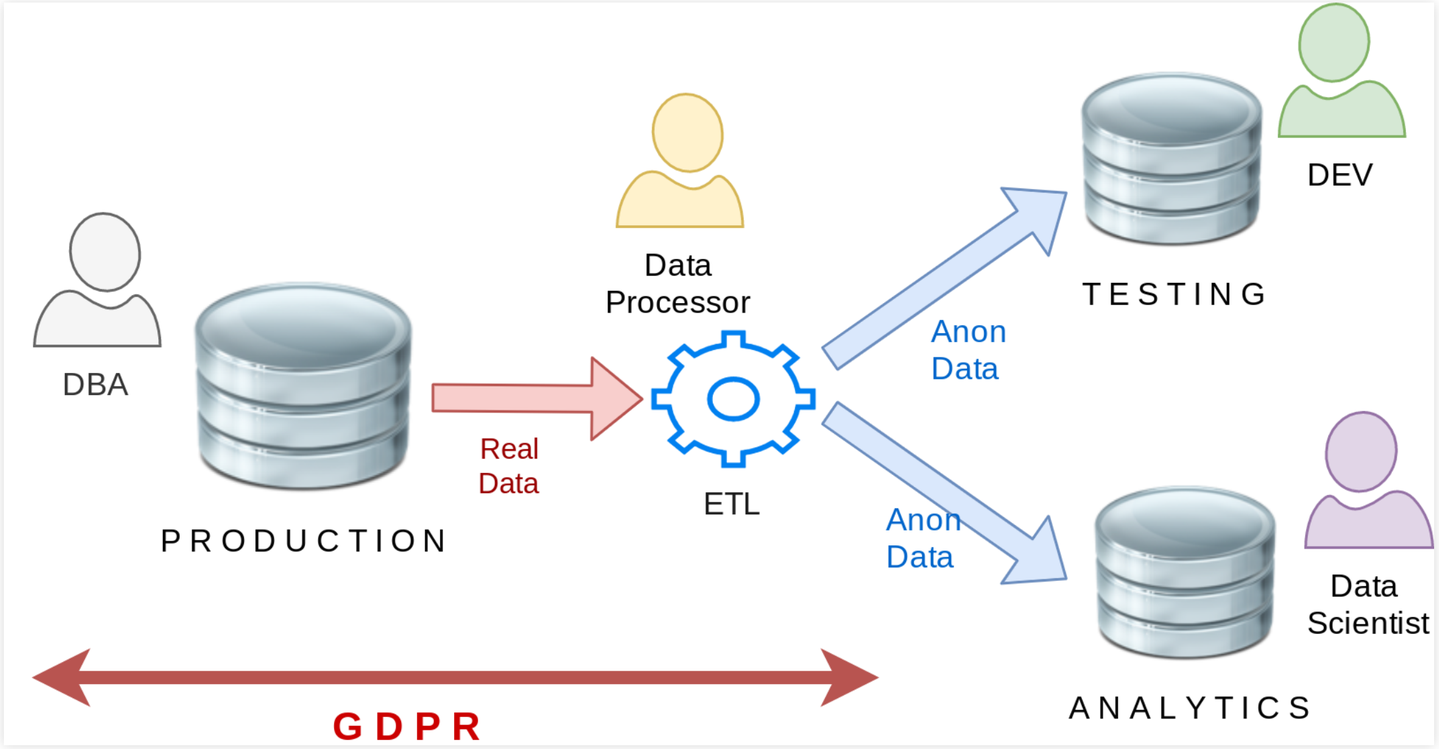
QUALIFIER LES ROLES

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

ANONYMISER DANS LA BASE

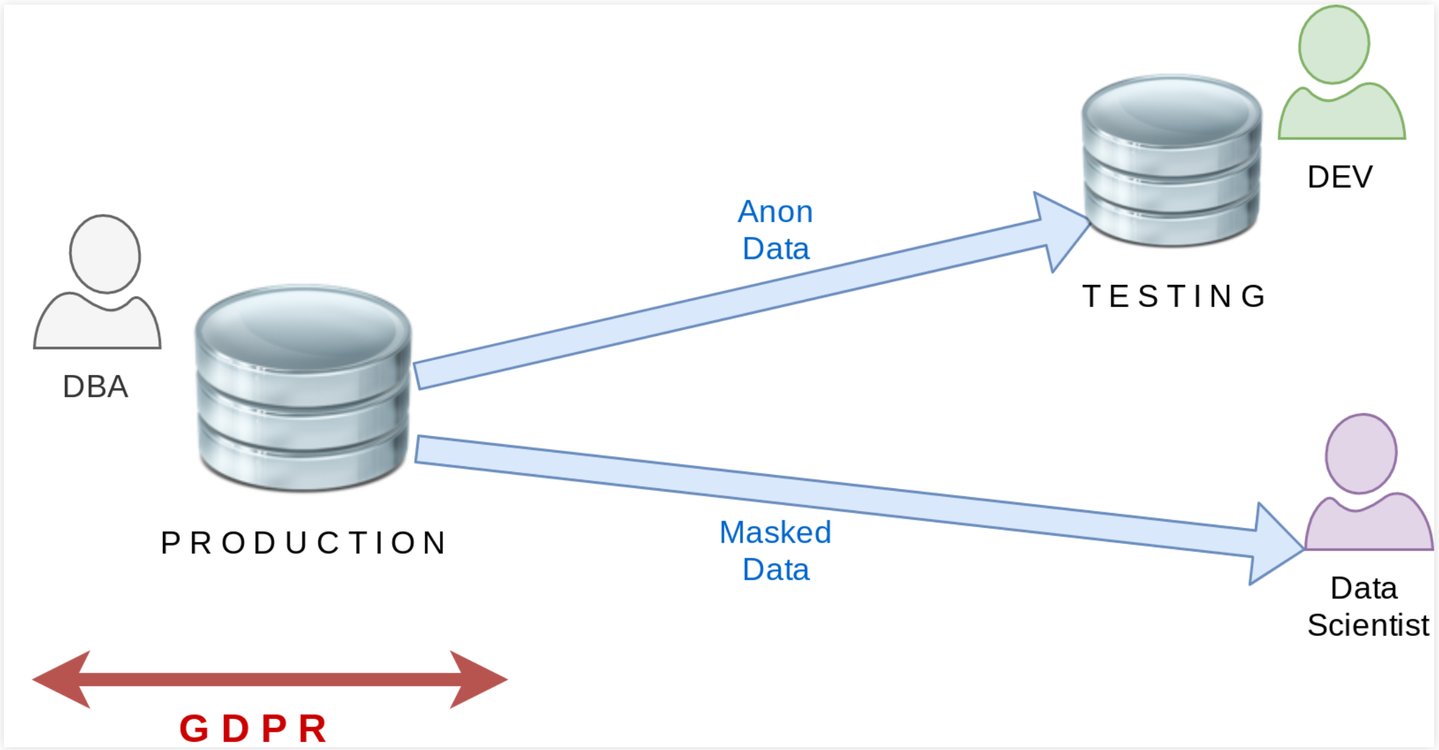


ANONYMISER DANS LA BASE



DALIBO
L'expertise PostgreSQL

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
- Video

COMMENT CONTRIBUER ?

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

https://gitlab.com/dalibo/postgresql_anonymizer

MERCI !

DGFIP

Biomerieux

Mes collègues

A BIENTÔT !

- Contact : damien.clochard@dalibo.com
- Follow : @daamien
- Nos autres Projets : Dalibo Labs