



**DU SIG À UNE BASE DE DONNÉES DE
RÉFÉRENCE DÉCISIONNELLE, QUELLE**

**ARCHITECTURE DE BASE DE
DONNÉES ?**

Comment survivre à l'anarchie sur le long terme?



/ME

Régis Haubourg

Géomaticien passionné

ex-administrateur de données géographiques pour une agence de l'eau

>>> "SQL saved my life"

Expert géomaticien Open Source chez Oslandia

QGIS & PostGIS fan

Président de l'OSGeo-FR



OSLANDIA

PME Française créée en 2009

Haute technologie

Open source

Systèmes d'Information Géographique

3D

DATA/IA



OSLANDIA

15 collaborateurs

100% télétravail

Entreprise horizontale et transparente

CA 2017 > 1.3 M€ / 30 % croissance / 20 % CA en R&D

100 % indépendant Projets 20-300K€ (→ 800K€ collab.
)



COMMUNAUTÉS OPEN SOURCE

QGIS

PostGIS

iTowns

python, SQL, C++, javascript

stack OSGEO - mais pas trop le côté JAVA ;-)



ET VOUS ?

Qui est arrivé à PostgreSQL par la géomatique et PostGIS?

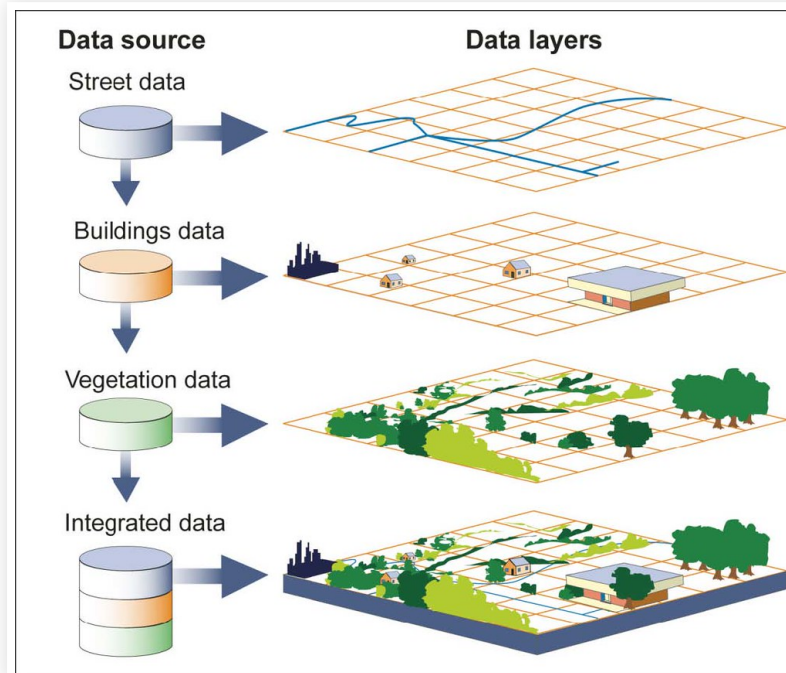
Votre base est-elle?:

- applicative ?
- décisionnelle (SIG / datascience / reporting) ?
- un peu des deux?



C'EST QUOI UN SIG ?

Systeme d'Information Géographique





C'ÉTAIT QUOI UN SIG ?

ArcGIS / Mapinfo / Geoconcept etc..

Un outil bureautique + des formats de stockage (fichier)
propriétaires

Un monde séparé de l'informatique traditionnelle

Un métier d'avenir! Mais personne ne comprends ce que
c'est.

*toute ressemblance avec le métier de datascientiste est
involontaire*



version géographe:

*"Représenter le monde, le comprendre, aider à la
décision, .. "*



Version commanditaire:

"Un truc pour imprimer des cartes"



Version DSI:

*"Encore un truc avec des serveurs de licences
ingérables et qui coûte trop cher"*



version DBA:

*"Une base de donnée avec des types de géométries
dedans quoi "*

.. ..

et la vie devint plus simple...



... enfin presque



NIVEAUX DE MATURITÉS DE LA DONNÉE DANS UNE ORGANISATION

Pas d'organisation de l'information numérique. Pas de conscience de l'utilité. Les applications priment sur la donnée



NIVEAUX DE MATURITÉS DE LA DONNÉE DANS UNE ORGANISATION

Reconnaissance de la valeur intrinsèque des données.
Collecte, catalogage, Entrepôt garage. "Open Data
washing"



NIVEAUX DE MATURITÉS DE LA DONNÉE DANS UNE ORGANISATION

Les données de référence sont structurées, centralisées, modélisées, harmonisées, et accessibles aux applications.

Un socle, mais beaucoup de données grises



NIVEAUX DE MATURITÉS DE LA DONNÉE DANS UNE ORGANISATION

Les projets applicatifs sont vus dès le départ sous l'angle
de la donnée et sa valeur. Les données métiers sont
standardisée, diffusées

Maturité



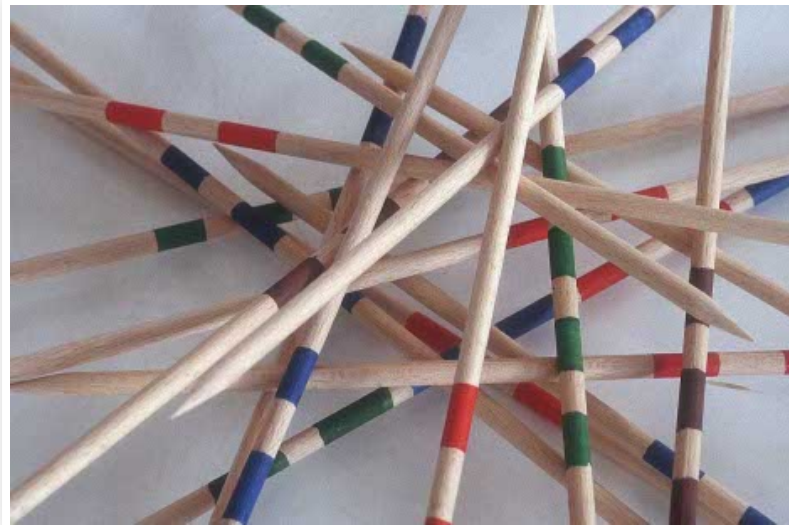
NIVEAUX DE MATURITÉS DE LA DONNÉE DANS UNE ORGANISATION

La valeur de la donnée est au centre dans le modèle de fonctionnement et génère directement de la valeur.

les licornes et GAFAM



LES SYMPTÔMES DU CHAOS





UNE BASE FIGÉE PAR LES APPLICATIONS CLIENTES





DES SCHÉMAS AVEC DES NOMS DE SERVICES DEDANS

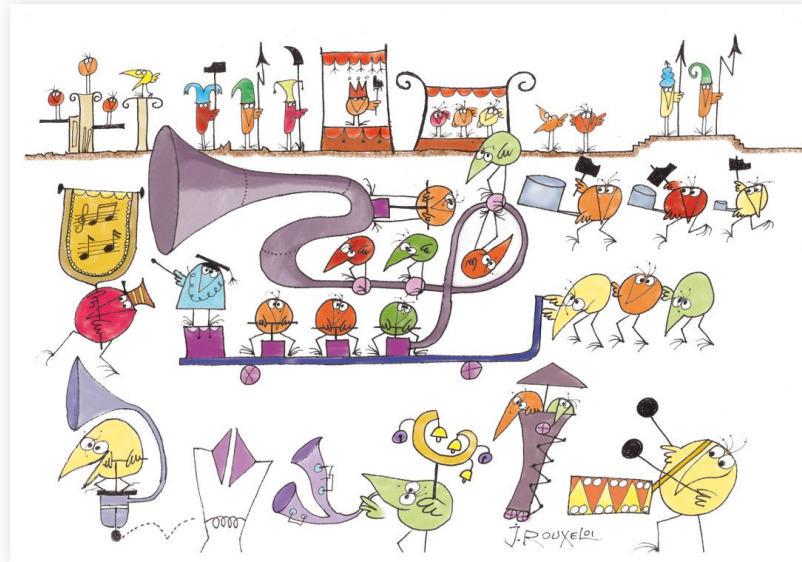
QGIS drag&drop vers PostGIS...

```
SELECT * FROM "serviceRPGD"."_Communes à facturer final1 " ;
```





DES ETLs TROP NOMBREUX ?





DES RÉPERTOIRES UTILISATEURS PLUS LOURDS QUE LA BASE DE DONNÉES

Agence de l'eau AG, la base de référence:
99% des données = 4 Go

5 tables = 15 Go

Chaque répertoire de fichier de référent SIG = **20Go !**



WHAT COULD POSSIBLY GO WRONG?

Ex dans une Métropole.

2016 : 70 schémas

2019 : 170 schémas

Des schémas avec des noms de services (acronymes)

Des fondations, c'est mieux si c'est stable..



COMMENT SURVIVRE ?



pour 20 à 50 ans..



WAIT..

Pourquoi les SIG sont-ils différents d'autres applications?

- Ils accèdent DIRECTEMENT à la donnée. Pas d'ORM
- Les données ne sont pas spécifiques à une application. 99% de données externes et de référentiels
- 90% des données sont potentiellement géographiques. On veut TOUT.



LES FONDATIONS



NOMMEZ LES CHOSES

```
SELECT
```

```
    zovacxov_nconnx,  
    trim(zovacxov_r_tyac)::text as zovacxov_r_tyac, --Information [co  
    zovacxov_r_pamo_cmatog ,  
    zovacxov_r_porc_cmatog::integer
```

```
FROM myenterprisename.zocxov
```



NOMMEZ LES CHOSES

Soyez concis, mais explicite

```
ref.parcelles  
ref.insee_recensement  
ref.admin_departement  
ref.routier_troncon  
ref.hydro_cours_eau  
ref.hydro_lac  
zon.zone_vulnerable
```



EXPLOITER LES VUES:

- Interface avec le modèle privé (tables)
- Remise en forme (le SIG aime les données en colonnes)
- Doc utilisateur

```
SELECT * from ref.v_admin_communes_recensement_insee2019
```



GARDER UN MODÈLE NON VERROUILLÉ

- Les vues sont éditables,
- faites en des abstractions entre le modèle de données privé et les services externes



LES PIÈGES !

Evitez les vues en cascade (1 à 2 niveaux max) :
verrouillage du modèle

Evitez au maximum les vues matérialisées

Analyser le plan d'exécution systématiquement /
optimiser

Les vues n'ont pas toutes métadonnées technique
(extent, clé primaire). Lenteur potentielle à l'ouverture



EMBRASSEZ LE MODÈLE ORIENTÉ SERVICE

ORM, API : OUI! (mais pas en SIG)

Gardez la logique métier dans la base

Vues simples

Procédures stockées pour remplir les API



VERSIONNEZ !

```
pg_dump --exclude-schema=public --schema-only my_precious_schema.sql  
git add my_precious_schema.sql  
git add my_previous_etl_loading_script.wtf  
git commit -m "my daily parachute"  
git push
```



UTTEZ CONTRE LES DONNÉES GRISES

Répondez aux besoins des utilisateurs

Formez au SQL

Formez au SQL

Formez au SQL (repeat after me)

Et aussi, tendre vers des traitements reproductibles
partout (script SQL, notebook, job ETL)



RTFM

Pour les administrateurs de données

```
COMMENT ON TABLE qwat_od.consumptionzone IS 'consumption zones. load
```

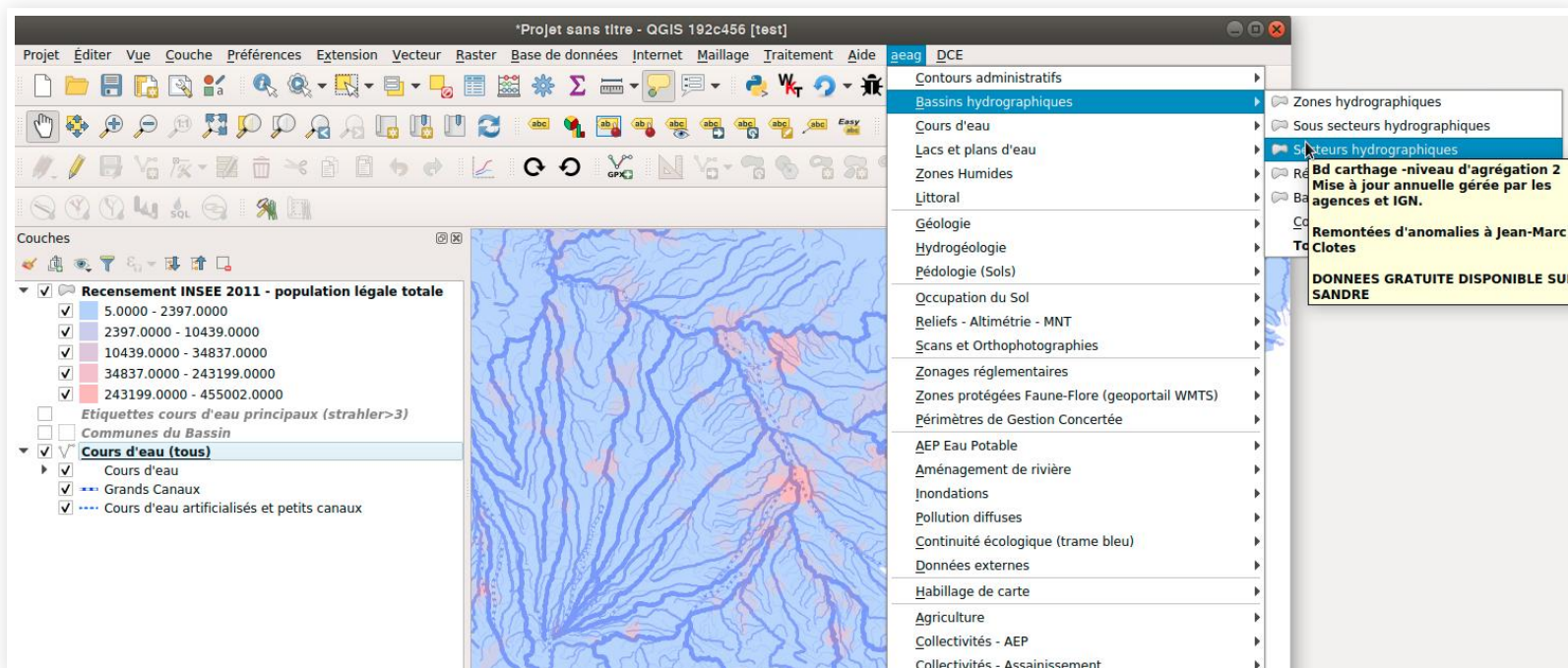
Pour les utilisateurs

```
COMMENT ON VIEW qwat_od.consumptionzone IS 'consumption zones. Repres
```



RTFCATALOG

Les utilisateurs veulent trouver ça:





CATALOGUER

Générez **automatiquement** des catalogues de métadonnées sur les informations orientées utilisateurs
(Ex: Talend Metadata crawler)

- couches SIG travaillées
- indexer des projets SIG et pas des bases (j'exagère)

Parecequ'on est pas des Shadoks



**NE METTEZ PAS FORCÉMENT
TOUT DANS LA BASE**





NE METTEZ PAS FORCÉMENT TOUT DANS LA BASE

GeoPackage will save our souls

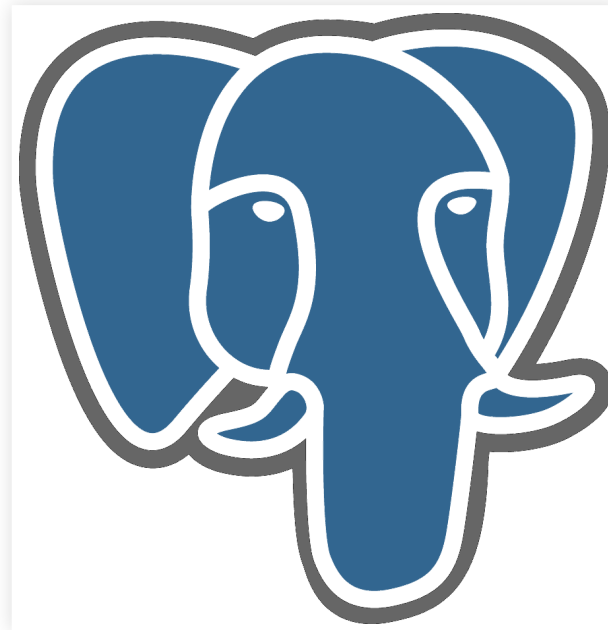


(hihi,c'est une base de données - basée sur SQLite /



NE METTEZ PAS FORCÉMENT TOUT DANS LA BASE

Un schéma PG bac à sable par utilisateur ?





EPAISSE MA BASE ?

Fournir l'intelligence métier au plus près de la donnée ?

Oui dans le cas d'une base SIG, multiclients



MONITOREZ

pgbadger, pgclue, powa

- vue sur l'usage réel de la base
- index et contraintes manquantes
- aspects techniques et performances



LE SERVICE SIG, PREMIER FOURNISSEUR DE RÉFÉRENTIELS

Charger des référentiels au quotidien. Rechercher la qualité. Diffuser. Utiliser.

Vers un service fournisseur de référentiel à toute l'organisation, au delà du SIG ?

Mais qui ne vient pas de l'informatique en général...



UN TRAVAIL SANS FIN





ET DONC DES HOMMES

Recrutez des administrateurs de données

Rapprochez administrateurs de données et
administrateurs de bases de données

Restez proche des utilisateurs

Faite comprendre les enjeux de la donnée dans chacun
des services

**MERCI DE VOTRE
ATTENTION**