

# pgQuaDoGéo

Qualité des données ? Ne vous y trompez pas, adoptez 2 éléphants !





# Avant propos



## Fiche d'identité

95%



5%

- Prénom: Michaël
- Nbm: Galien
- Surnom: Tetranos
- Age : 39 ans
- Loisirs: Running, Electro, Supporter **USAM**
- Profession : ginformaticien
- Employeur : Gard
- E-mail : michael.galien@gard.fr
- Bluesky : @tetranos.bsky.social
- Mastodon : @tetranos@mapstodon.space
- Twitter : @tetranos

# Remerciements



**IDEO**

BOURSCOVNE - FRANCHÉ - LOCALITÉ



**ARNiA**

Agence régionale du numérique  
et de l'intelligence artificielle



le Crige



**OPENIG**



**Cerema**



**ARNiA**

Agence régionale du numérique  
et de l'intelligence artificielle



**IDEO**

BOURSCOVNE - FRANCHÉ - LOCALITÉ

# Si vous avez manqué le début...

2013 - Norme ISO 19157 sur la qualité des données



2017 - 11 fiches CEREMA de décryptage de la norme 19157

<https://www.cerema.fr/fr/actualites/serie-fiches-cerema-qualifier-donnees-geographiques>



2021/2022/2023 - Sessions QuaDoGéo (QGIS & PostgreSQL/PostGIS)  
dans le cadre des travaux du GT QuaDoGéo du CNG  
co-organisées par le CRIGE PACA, le CEREMA et OPenIG



2023 - On prend les mêmes et on recommence  
avec l'aide d'IdeoBFC et de l'ARNA



BOURBONNE - FRANCE - 04 - 100119



**ARNiA**  
Agence régionale du numérique  
et de l'intelligence artificielle



# Programme

- 14/11/2023 - Session 1 : Modélisation et contrôle a priori
- 21/11/2023 - Session 2 : Règles assouplies et contrôle a posteriori
- 28/11/2023 - Session 3 : (Géo)analyses avancées

# Déroulé

- Les sessions alterneront théorie (PowerPoint) et pratique (SQL).
- Les scripts seront exécutés sous pgAdmin et, si géographiques, les résultats seront visualisés avec QGIS.
- Les scripts SQL ont été prérédigés :
  - moins de risques « effet démo »,
  - plus de temps pour les commenter.
- Le PowerPoint et les scripts SQL seront mis à disposition à l'issu des présentations sous licence CC-BY.

# pgQuaDoGéo : Session 1

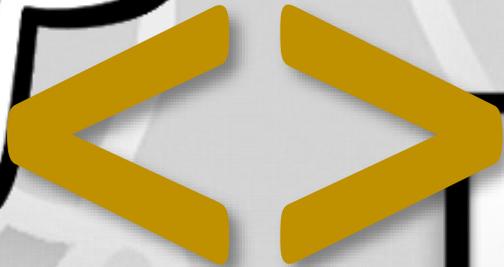
Modélisation et contrôle a priori

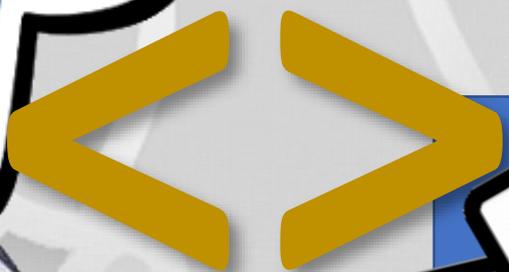




# Quand un SGBDg rencontre un besoin







# define:sgbd src:wikipedia.fr

Un SGBD permet d'enregistrer des données, puis de les rechercher, de les modifier et de créer automatiquement des comptes rendus (anglais report) du contenu de la base de données. Il permet de spécifier les types de données, la structure des données contenues dans la base de données, ainsi que des règles de cohérence telles que l'absence de redondance.

Source :

[https://fr.wikipedia.org/wiki/Syst%C3%A8me\\_de\\_gestion\\_de\\_base\\_de\\_donn%C3%A9es](https://fr.wikipedia.org/wiki/Syst%C3%A8me_de_gestion_de_base_de_donn%C3%A9es)

# Les expressions de la qualité

- La **qualité interne** : elle exprime la qualité d'un lot de données produit au regard de nécessaires spécifications, rédigées au préalable pour décrire précisément ce qui doit être produit (cahier des charges).
- La **qualité externe** : elle se définit plutôt comme l'adéquation à un besoin exprimé. Dit différemment, c'est l'aptitude d'un jeu de données à satisfaire un usage donné. Il faut bien entendre ici un type d'usage générique comme une exploitation cartographique, un usage à but statistique, ou la constitution d'un observatoire.
- La **qualité perçue** : elle exprime la perception d'un utilisateur pour un besoin spécifique. Cette approche (...) a été démocratisée avec l'essor de l'Internet et la multiplication des systèmes de notation par les communautés d'internautes.

Source :

Fiche 2 de décryptage de la norme ISO 19157 par le CEREMA

# Qualité implique besoin et vice-versa

- La qualité s'apprécie au regard d'un besoin ou d'une spécification (donc d'un besoin).
  - Adéquation précision/besoin :  $\pi \approx 3,1415926535897932384626433832795028841971693993751$
  - Adéquation format/besoin :  $\pi \approx \text{CCCIV} / \text{C}$
  - Adéquation support/besoin : « Veuillez trouver ci-joint le PDF contenant la valeur de  $\pi$  à 1000 décimales. »

## Fiche d'identité

95%



METHODE D'ESTIMATION UTILISEE ??!?



5%

- Prénom: Michaël
- Nbm: Galien
- Surnom: Tetranos
- ~~Age: 39 ans~~ *FAUX, ou VRAI à +/- 5 ans*
- Loisirs: Running, Electro, Supporter **USAM** *HORS SUJET*
- Profession: informaticien *HORS DU DOMAINE DE VALEURS*
- Employeur: Gard
- E-mail: michael.galien@gard.fr
- Bluesky: @tetranos.bsky.social
- Mastodon: @tetranos@mapstodon.space
- Twitter: @tetranos

# Modéliser = ST\_Simplify(le\_monde)

- Un SGBD est capable d'assurer la cohérence des données...
- ...encore faut-il lui dire quelles sont les règles.
  
- Modéliser (ou concevoir) :
  - c'est faire une abstraction du monde réel (c-à-d décrire « sa » vision du monde),
  - pour aboutir à un modèle de données qui répond à ses besoins et contraintes.

# fonction modeliser(mes\_besoins) returns mon\_monde {

1. identifier\_entités()
2. lister\_attributs(entités)
3. définir\_formats\_et\_valeurspossibles(attributs)
4. établir\_relations(entités)
5. normaliser()
6. si (bénéfice > risque) { dénormaliser\_partiellement() }

}

# Exemple RH – Entités, attributs et types

## POSTES

Nom: Chaîne

## SERVICES

Nom: Chaîne

## AGENTS

Nom: Chaîne  
Prénom: Chaîne  
Genre: F/M  
E-mail: Chaîne

## ADRESSES

Numéro: Entier  
Voie: Chaîne  
Commune: Chaîne

b. To which gender identity do you most identify?

**A** Female

**B** Male

**C** Transgender Female

**D** Transgender Male

**E** Gender Variant/Non-Conforming

**F** I do not wish to answer this question

**G** Other

OK ✓



Vous êtes :

Un homme ✓

Non-binaire

Une femme

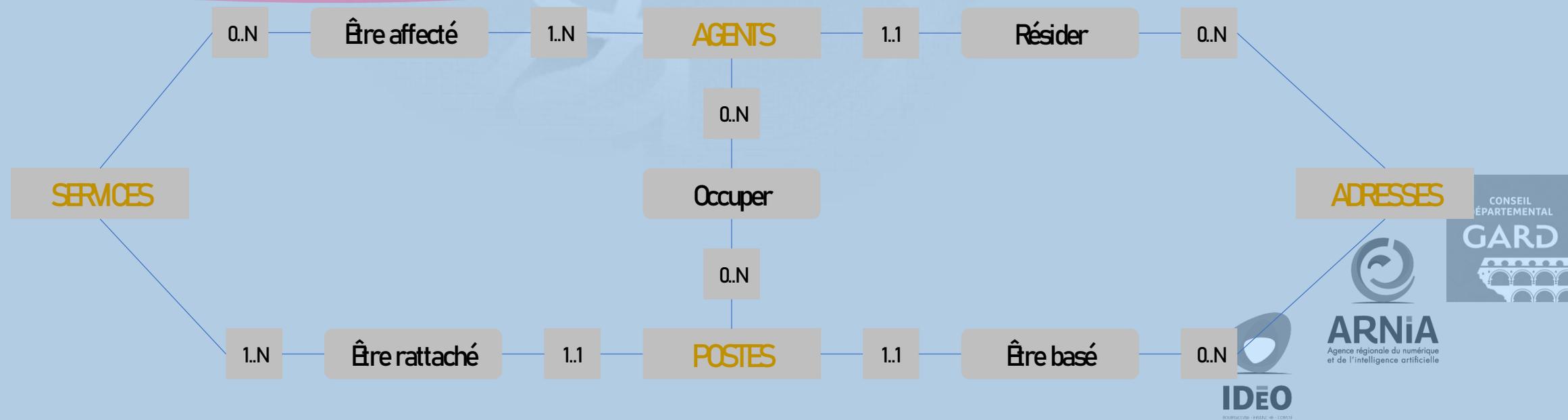
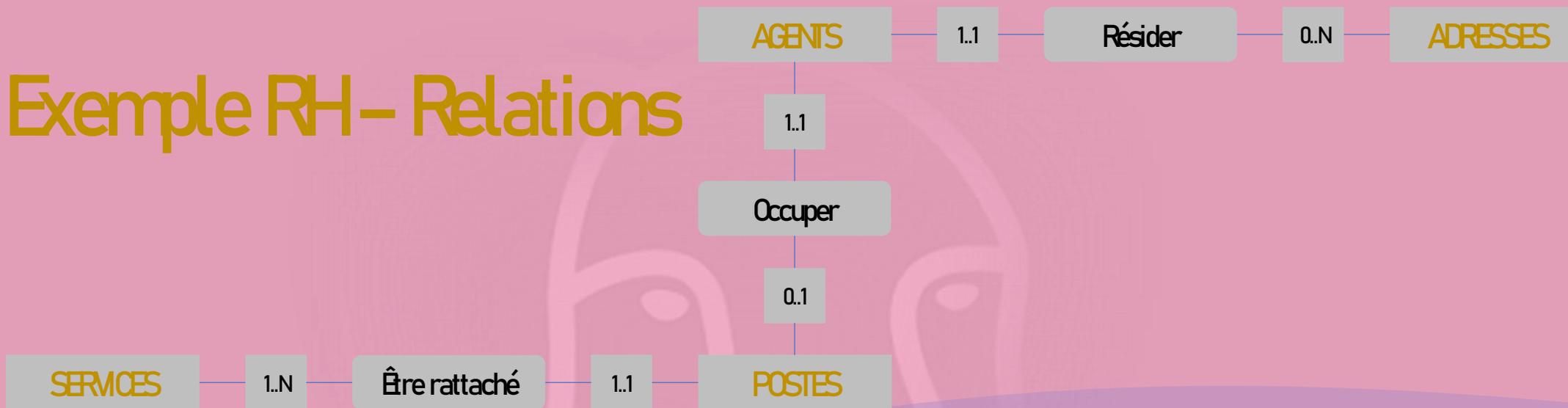
Ne souhaite pas répondre



# Exemple RH – Relations

- Un agent peut-il occuper plusieurs postes ?
  - **Nbn** VS **Oui** (2 mi-temps sur des affectations différentes)
- Un agent peut-il ne pas avoir de poste ?
  - **Nbn** VS **Oui** (un stagiaire n'occupe pas un poste)
- Un poste peut-il être occupé par plusieurs agents ?
  - **Nbn** VS **Oui** (un agent en décharge syndicale, poste complété par un mi-temps)
- Un agent peut-il être rattaché à un service différent de celui de son poste ?
  - **Nbn** VS **Oui** (cas des reconversions professionnelles)
- Avez-vous plusieurs sites ?
  - **Nbn** VS **Oui**

# Exemple RH – Relations





pgQuaDoGéo - 1 - Modèle non normalisé



# Les formes normales



**ARNiA**  
Agence régionale du numérique  
et de l'intelligence artificielle



# define:formes normales src:wikipedia.fr

La normalisation consiste à restructurer une base de données pour respecter certaines formes normales, afin d'éviter la redondance des données (des données apparaissent plusieurs fois) et d'assurer l'intégrité des données.

Source :

[https://fr.wikipedia.org/wiki/Forme\\_normale\\_\(bases\\_de\\_donn%C3%A9es\\_relationnelles\)](https://fr.wikipedia.org/wiki/Forme_normale_(bases_de_donn%C3%A9es_relationnelles))

Un attribut ne doit pas être dépendant d'une sous-partie de la clé dès..

♦ A: La 1NF

♦ B: La 2NF

♦ C: La 3NF

♦ D: La BCNF

Une entité est en 1NF si elle possède au moins une clé et si tous ses attributs sont atomiques.

Un attribut est atomique s'il ne contient qu'une seule valeur pour un tuple donné, et donc s'il ne regroupe pas un ensemble de plusieurs valeurs.

1NF

2NF

3NF

BCNF

Une entité est en 3NF si elle est en 2NF et si tout attribut n'appartenant pas à une clé ne dépend pas d'un autre attribut n'appartenant pas à une clé. C'est à dire encore que toutes les dépendances fonctionnelles élémentaires vers des attributs n'appartenant pas à une clé, sont issues d'une clé.

Une entité est en BCNF si elle est en 3NF et si tout attribut qui n'appartient pas à une clé n'est pas source d'une dépendance fonctionnelle vers une partie d'une clé. C'est à dire que les seules dépendances fonctionnelles élémentaires existantes sont celles dans lesquelles une clé détermine un attribut.

# Contre-exemple 1NF: absence de clé

Table: AGENT

nom

prenom

genre

date\_naissance

email

type\_emploi

id\_poste [FK]

id\_adresse [FK]

# Contre-exemple 1NF: absence de clé

Table: AGENT  
nom  
prenom  
genre  
date\_naissance

Query Query History

```
1 select * from agent
```

Data Output Messages Notifications

	nom character varying	prenom character varying	genre character varying (1)	date_naissance date	email character varying	type_emploi character varying	id_poste integer	id_adresse character varying
1	CHARDAC	CELINE	F	1969-09-06	celine.chardac@pgquadogeo.fr	contractuel	1989	30138_0078_00633_bis
2	CHARDAC	CELINE	F	1969-09-06	celine.chardac@pgquadogeo.fr	contractuel	1989	30138_0078_00633_bis
3	GAUBERT	JULIEN	M	1972-03-23	julien.gaubert@pgquadogeo.fr	titulaire	2616	30061_0245_00002_c
4	BEKAKCHA	CHANTAL	F	1986-09-10	chantal.bekakcha@pgquadogeo.fr	titulaire	2608	30145_0250_00009_bis

# Contre-exemple 1NF: attribut non atomique

Table : ADRESSE
id [PK]
libelle
code_postal_commune [FK]
nom_commune [FK]

# Contre-exemple 1NF: attribut non atomique

Table: ADRESSE  
id [PK]  
libelle  
code\_postal\_commune [FK]

Query Query History

```
1 select * from adresse
```

Data Output Messages Notifications

	id [PK] character varying	libelle character varying	code_postal_commune character varying	nom_commune character varying	geom geometry
1	30189_3055_00001	1 Rue Faraday	30000	Nimes	01010000206A0800008FC2F5A8EE9D284100000020650C5841
2	30039_0157_00031	31 Rue de la Tapie	30320	Bezouze	01010000206A0800003D0AD7A31E012941D7A370FD511258...
3	30317_0055_00004	4 Grand Rue	30210	Sernhac	01010000206A08000052B81E05812A294114AE4701B3155841
4	30122_0041_00039	39 Grand Rue	30730	Gajan	01010000206A080000EC51B89E7657284114AE4751B5135841
5	30258_0022_00001_bis	1 bis Rue Antoine Pages	30800	Saint-Gilles	01010000206A080000A4703D0A09E62841CDCCCCBC27FD5...
6	30258_0023_00005	5 Impasse des Andalous	30800	Saint-Gilles	01010000206A08000066666666B8E72841D7A3701DDB0058...

# Contre-exemple 2NF: code\_postal > cog\_departement

Table: COMMUNE

code\_postal [PK]

nom [PK]

cog

cog\_departement [FK]

# Contre-exemple 2NF: code\_postal > cog\_departement

Table: COMMUNE  
code\_postal [PK]  
nom [PK]

Query Query History

```
1 select * from commune where code_postal = '30190'
```

Data Output Messages Notifications

	code_postal [PK] character varying (5)	nom [PK] character varying	cog character varying (5)	cog_departement character varying (3)	geom geometry
1	30190	Sainte-Anastasie	30228	34	01060000206A080000010000000103000
2	30190	Saint-Genies-de-Malgoires	30255	30	01060000206A080000010000000103000
3	30190	Saint-Chaptes	30241	30	01060000206A080000010000000103000
4	30190	La Calmette	30061	30	01060000206A080000010000000103000
5	30190	Sauzet	30313	30	01060000206A080000010000000103000
6	30190	Montignargues	30180	30	01060000206A080000010000000103000
7	30190	La Rouviere	30224	30	01060000206A080000010000000103000
8	30190	Dions	30102	30	01060000206A080000010000000103000



ARNiA  
Service régional de numérique  
à l'intelligence artificielle



# Contre-exemple 3NF: cog\_region > nom\_region

Table: DEPARTEMENT

cog [PK]

nom

cog\_region

nom\_region

# Contre-exemple 3NF: cog\_region > nom\_region

Table: DEPARTEMENT

cog [PK]

nom

cog\_region

nom\_region

Query Query History

```
1 select * from departement
```

Data Output Messages Notifications

	cog [PK] character varying (3)	nom character varying	cog_region character varying (2)	nom_region character varying
1	07	Ardeche	84	Auvergne-Rhone-Alpes
2	48	Lozere	76	Occitanie
3	12	Aveyron	76	Ocitanie
4	84	Vaucluse	93	Provence-Alpes-Cote d'Azur

# Contre-exemple BCNF : cog > nom

Table: COMMUNE

code\_postal [PK]

nom [PK]

cog

cog\_departement [FK]

# Contre-exemple BCNF : cog > nom

Table: COMMUNE  
code\_postal [PK]  
nom [PK]  
cog  
cog\_departement [FK]

Query Query History

```
1 select * from commune where code_postal = '30320'
```

Data Output Messages Notifications

	code_postal [PK] character varying (5)	nom [PK] character varying	cog character varying (5)	cog_departement character varying (3)
1	30320	Bezouce	30039	30
2	30320	Saint-Gervasy	30257	30
3	30320	Marguerittes	30156	30
4	30320	Poulx	30206	30

# En résumé

- 1NF: ko si absence de clé ou champ non atomique.
- 2NF: ko si une sous-partie de la clé → attribut.
- 3NF: ko si un attribut non clé → un attribut non clé.
- BCNF: ko si un attribut non clé → une sous-partie de la clé.

Un attribut ne doit pas être dépendant d'une sous-partie de la clé dès..

♦ A: La 1NF

♦ B: La 2NF

♦ C: La 3NF

♦ D: La BCNF



## pgQuaDoGéo - 2 - Normalisation

# Géostandards et schémas





# Schémas : Ex. Arbres urbains

exemple-valide.xlsx - Excel

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Aide Rechercher des outils adaptés Partager

Coller Presse-papiers Police Alignement Nombre Styles Cellules Édition

A4 : 3

	A	B	C	D	E	F	G	H	I	J	K	
1	id	longitude	latitude	date_releve	famille	genre	espece	cultivar_vari	nom_vernac	code_insee	code_postal	adresse
2	1	49.802	0.3768	2018-10-12	Sapindaceae	Acer	campestre		Erable champ	2A123	13600	320 Avenue Joseph I
3	2	43.2959	5.3761	2021-04-20	Rosaceae	Pyrus	alleryana	chanticleerc	Poirier "char	75013	13001	60 rue sainte
4	3	43.2915	5.4805	1998-03-01	Platanaceae	Platanus	occidentalis		Platane d'An	13112	13002	72 Boulevard des Da
5	4	44.2920	5.3705		Pinaceae	Pinus	sylvestris		Pin sylvestre	14654	13013	48 rue des hauts boi
6												
7												
8												

Prêt Accessibilité : vérification terminée

115 %

# Schémas : Ex. BAL

CONTRAIREA3NF

CONTRAIREA1NF

attribut	obligatoire	description
id_ban_commune		identifiant unique de la commune fourni par la BAN
id_ban_toponyme		identifiant unique du toponyme auquel est rattaché l'adresse
id_ban_adresse		identifiant unique de l'adresse
cle_interop	X	clé nationale d'interopérabilité
commune_insee	X	code INSEE de la commune
commune_nom	X	nom de la commune
commune_deleguee_insee		code INSEE de la commune déléguée
commune_deleguee_nom		nom de la commune déléguée
voie_nom	X	nom complet de la voie
lieudit_complement_nom		nom du lieu-dit historique ou complémentaire
numero	X	numéro
suffixe		informations qui complètent et précisent les numéros d'adresses
position	X	décrit la position d'une adresse (liste de valeurs)
x	X	système de projection légal en vigueur sur le territoire concerné
y	X	
long	X	coordonnées exprimées en WGS84
lat	X	
cad_parcelles		liste des parcelles, séparées par un pipe ( ), desservies ou représentées par cette adresse
source	X	organisme ayant créé cette adresse
date_der_maj	X	Date de la dernière mise à jour de la donnée au format AAAA-MM-JJ
certification_commune	X	Certification communale

# Et la BDTopo IGN?

Query Editor Query History

```
1 SELECT cleabs, insee_commune_gauche, code_postal_gauche, insee_commune_droite, code_postal_droit, cpx_gestionnaire, cpx_numero
2 FROM bdtopo_troncon_de_route
3 WHERE cpx_classement_administratif = 'Départementale'
4 AND cpx_numero LIKE '%/%'
5 LIMIT 100
```

« CONTRAIRE A 3NF »

CONTRAIRE A 1NF

Data Output Explain Messages Notifications

	cleabs character varying (24)	insee_commune_gauche character varying (5)	code_postal_gauche character varying (5)	insee_commune_droite character varying (5)	code_postal_droit character varying (5)	cpx_gestionnaire character varying	cpx_numero character varying (32)
1	TRONROUT0000000053381239	78172	78700	95218	95610	Val-d'Oise/Yvelines	D48E/D54
2	TRONROUT0000000075788776	05163	05160	05163	05160	Alpes-de-Haute-Provence/Hautes-Alpes	D954/D954
3	TRONROUT0000000075787807	05163	05160	05163	05160	Alpes-de-Haute-Provence/Hautes-Alpes	D954/D954
4	TRONROUT0000000053381236	78172	78700	95218	95610	Val-d'Oise/Yvelines	D48E/D54
5	TRONROUT0000000043882167	70155	70300	70155	70300	Haute-Saône/Haute-Saône	D64BC20/D64BD20
6	TRONROUT0000000227715171	05163	05160	05163	05160	Alpes-de-Haute-Provence/Hautes-Alpes	D954/D954
7	TRONROUT0000000075787804	05163	05160	05163	05160	Alpes-de-Haute-Provence/Hautes-Alpes	D954/D954
8	TRONROUT0000000352884119	05163	05160	05163	05160	Alpes-de-Haute-Provence/Hautes-Alpes	D954/D954



IDEO

ARNiA  
Agence régionale du numérique  
et de l'intelligence artificielle

BOURBONNE - FRANCE - 100110

# Où les trouver ?

- Standards CNG
  - <http://cnig.gouv.fr/les-standards-cnig-a18959.html>
- Standards COVADIS
  - <http://www.geoinformations.developpement-durable.gouv.fr/standards-covadis-valides-r469.html>
- Schémas Data.Gouv
  - <https://schema.data.gouv.fr/schemas.html>

# DDL (Data Definition Language)



# Les langages SQL

- DML (Data Manipulation Language)
  - SELECT
  - INSERT
  - UPDATE
  - DELETE
- DDL (Data Definition Language)
  - CREATE
  - ALTER
  - DROP
  - TRUNCATE
- DCL (Data Control Language)
  - GRANT
  - REVOKE
- TCL (Transaction Control Language)
  - COMMIT
  - ROLLBACK

# Démo

```
ALTER TABLE agent  
ADD CONSTRAINT agent_genre  
CHECK (genre IN ('M', 'F'));
```



## Fin de la session 1

### Réflexions

- Un modèle normalisé est un modèle ramifié, mal adapté au partage et à l'OpenData.
- En général, publier en OpenData c'est dénormaliser c'est-à-dire produire des exports à plat.
- Il est plus facile de dénormaliser que l'inverse.

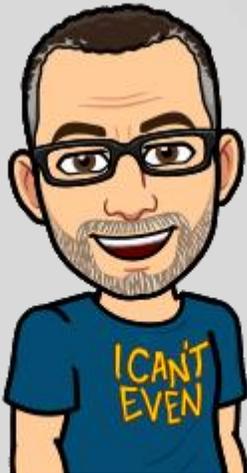


## Fin de la session 1

### Récap'

- La qualité s'apprécie au regard d'un besoin.
- Les contraintes/besoins peuvent être résolument différents d'un acteur à l'autre.
- Modéliser une base de données c'est décrire « sa » vision du monde.
- La mise en œuvre du modèle dans le SGBD permet d'imposer « by design » le respect de certaines contraintes.

MERC!



# Des questions ?



# pgQuaDoGéo : Session 2

Règles assouplies et contrôle a posteriori



## Fiche d'identité



- Prénom: Michaël
- Nbm: Galien
- Surnom: Tetranos
- Profession : Administrateur de la donnée
- Employeur : Gard
- E-mail : michael.galien@gard.fr
- Bluesky : @tetranos.bsky.social
- Mastodon : @tetranos@mapstodon.space
- Twitter : @tetranos

Si vous avez manqué  
la session 1...



- La qualité s'apprécie au regard d'un besoin.
- Les contraintes/besoins peuvent être résolument différents d'un acteur à l'autre.
- Modéliser une base de données c'est décrire « sa » vision du monde.
- La mise en œuvre du modèle dans le SGBD permet d'imposer « by design » le respect de certaines contraintes.

# Contraint oui, bloquant non !

- Nous avons vu en session 1 comment concevoir sa BDD pour assurer la cohérence des données :
  - via une modélisation normalisée et adaptée aux besoins/contraintes,
  - via différents « outils » proposés par le DDL
- Mais, on ne maîtrise pas toujours le cycle de vie de la donnée (data analyst vs data owner).
  - Données internes : qui corrige ? quand ?
  - Données externes : fréquence des mises à jour ? signalements ?
- Il peut être utile/nécessaire de faire un modèle moins contraint.

# Moins contraint ? Comment ?

- Suppression des CHECK, TRIGGER, ...
- Autorisation des valeurs à NULL
- Typage faible VARCHAR/TEXT.
- Assouplissement des FOREIGNKEY, voire suppression.
- Dénormalisation.
- ...

# Moins contraint ? Exemple.



# Modèle non contraint bien connu

```
CREATE TABLE public.node_tags (  
  node_id bigint NOT NULL,  
  version bigint NOT NULL,  
  k character varying DEFAULT ''::character varying NOT NULL,  
  v character varying DEFAULT ''::character varying NOT NULL  
);  
  
CREATE TABLE public.nodes (  
  node_id bigint NOT NULL,  
  latitude integer NOT NULL,  
  longitude integer NOT NULL,  
  changeset_id bigint NOT NULL,  
  visible boolean NOT NULL,  
  "timestamp" timestamp without time zone NOT NULL,  
  tile bigint NOT NULL,  
  version bigint NOT NULL,  
  redaction_id integer  
);  
  
ALTER TABLE ONLY public.node_tags  
  ADD CONSTRAINT node_tags_pkey PRIMARY KEY (node_id, version, k);  
  
ALTER TABLE ONLY public.nodes  
  ADD CONSTRAINT nodes_pkey PRIMARY KEY (node_id, version);  
  
ALTER TABLE ONLY public.node_tags  
  ADD CONSTRAINT node_tags_id_fkey FOREIGN KEY (node_id, version) REFERENCES public.nodes(node_id, version);
```

# Et la qualité alors ?

Elle peut être vérifiée/mesurée par l'exécution de multiples requêtes (DML - DATA MANIPULATION LANGUAGE) sur les données pour :

- vérifier la cohérence des relations entre entités,
- vérifier/mesurer la qualité des attributs qualitatifs (taux de conformité),
- vérifier/mesurer la qualité des attributs quantitatifs (statistiques basiques),
- vérifier/mesurer la qualité d'un jeu de données vis-à-vis d'un référentiel (taux d'exhaustivité, précision absolue, précision relative),
- produire des échantillons.



pgQuaDoGéo - 4 - Préparation session 2.sql

# Démo

```
SELECT *  
FROM agent  
WHERE genre IS NULL  
OR genre NOT IN ('F', 'M');
```



# Dans la démo, nous avons vu comment :

- assouplir/supprimer les contraintes de référence et en vérifier la cohérence,
- vérifier la cohérence des attributs qualitatifs et calculer le taux de (non)conformité,
- calculer des statistiques basiques sur les attributs quantitatifs,
- identifier les excédents/omissions d'un jeu vis-à-vis d'un référentiel,
- calculer le taux d'exhaustivité d'un jeu vis-à-vis d'un référentiel,
- vérifier la précision géographique absolue et relative d'un jeu vis-à-vis d'un référentiel.

# Et nous avons vu comment faire des échantillons :

- simple aléatoire : 100 adresses choisies au hasard,
- orienté objet : toutes les adresses liées à une place,
- orienté surface : toutes les adresses d'une des communes,
- aléatoire stratifié : 5 adresses choisies au hasard dans chaque commune
- semi-aléatoire : calcul d'une grille sur l'emprise d'étude et sélection des adresses qui intersectent avec une cellule sur N

## Fin de la session 2



### Récap'

- Le contrôle a priori peut s'avérer « bloquant » si on n'est pas maître de la donnée d'entrée.
- Un modèle moins contraint, qui tolère l'ingestion de certaines erreurs, couplé à des requêtes d'analyse/détection de ces erreurs est une alternative plus souple.
- Différentes analyses a posteriori sont possibles :
  - Attribut qualitatif : taux de conformité,
  - Attribut quantitatif : étendue, moyenne, variance, écart-type,
  - Jeu : taux d'exhaustivité et précision vis-à-vis d'un référentiel.
- Plusieurs techniques permettent la création d'échantillons.

# Des questions ?



**ARNiA**  
Agence régionale du numérique  
et de l'intelligence artificielle



# pgjQuaDoGéo : Session 3

(Géo)analyses avancées



## Fiche d'identité



- Prénom: Michaël
- Nbm: Galien
- Surnom: Tetranos
- Profession : Administrateur de la donnée
- Employeur : Gard
- E-mail : michael.galien@gard.fr
- Bluesky : @tetranos.bsky.social
- Mastodon : @tetranos@mapstodon.space
- Twitter : @tetranos

Si vous avez manqué  
la session 2..



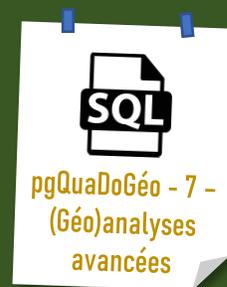
- Le contrôle a priori peut s'avérer « bloquant » si on n'est pas maître de la donnée d'entrée.
- Un modèle moins contraint, qui tolère l'ingestion de certaines erreurs, couplé à des requêtes d'analyse/détection de ces erreurs est une alternative plus souple.
- Différentes analyses a posteriori sont possibles :
  - Attribut qualitatif : taux de conformité,
  - Attribut quantitatif : étendue, moyenne, variance, écart-type,
  - Jeu : taux d'exhaustivité et précision vis-à-vis d'un référentiel.
- Plusieurs techniques permettent la création d'échantillons.



pgQuaDoGéo - 6 - Préparation session 3.sql  
+ import data\gpkg

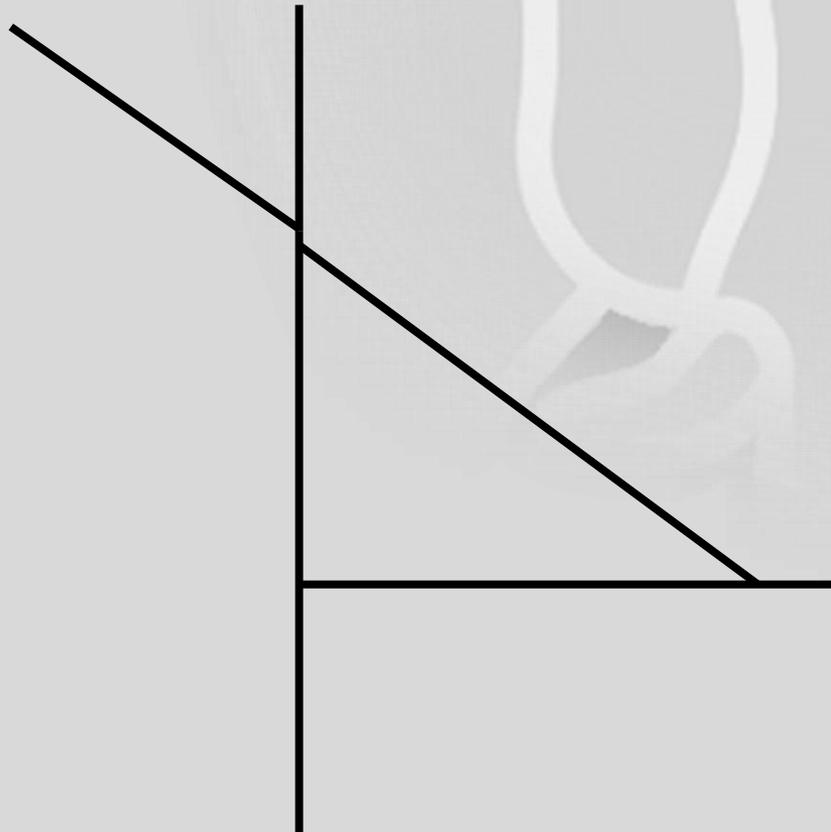
# Démo

```
SELECT ST_IsvalidReason(geom), *  
FROM commune  
WHERE NOT ST_Isvalid(geom);
```

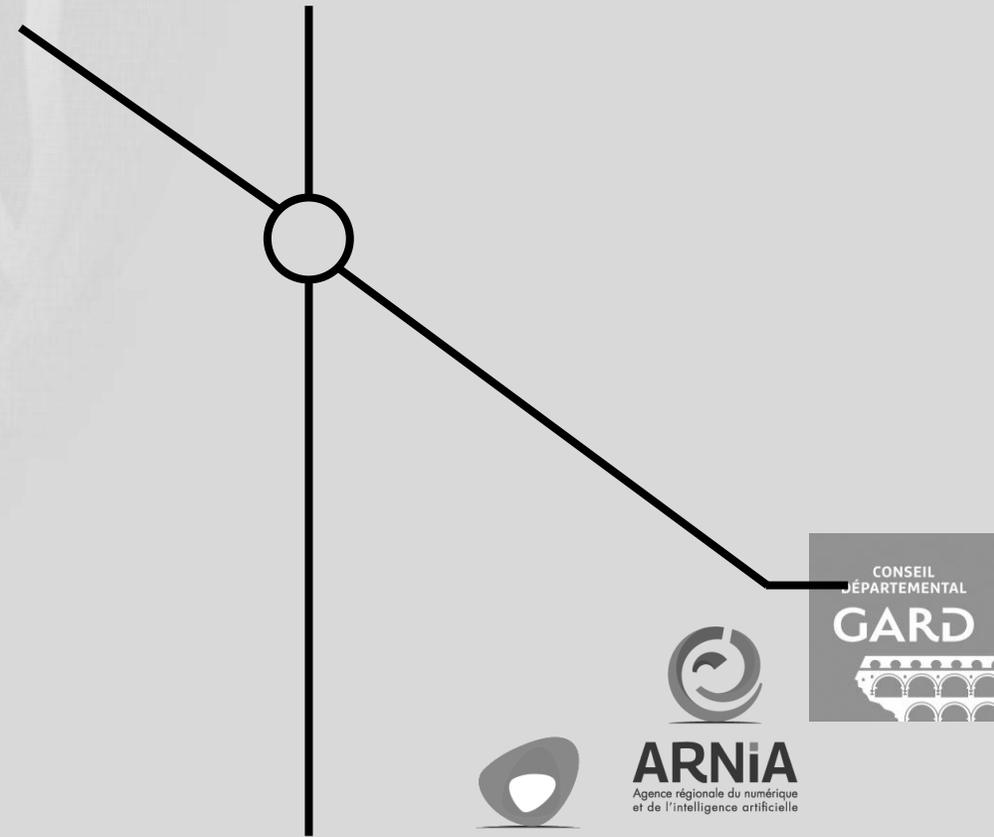


# Comparaison de 2 linéaires

Linéaire 1

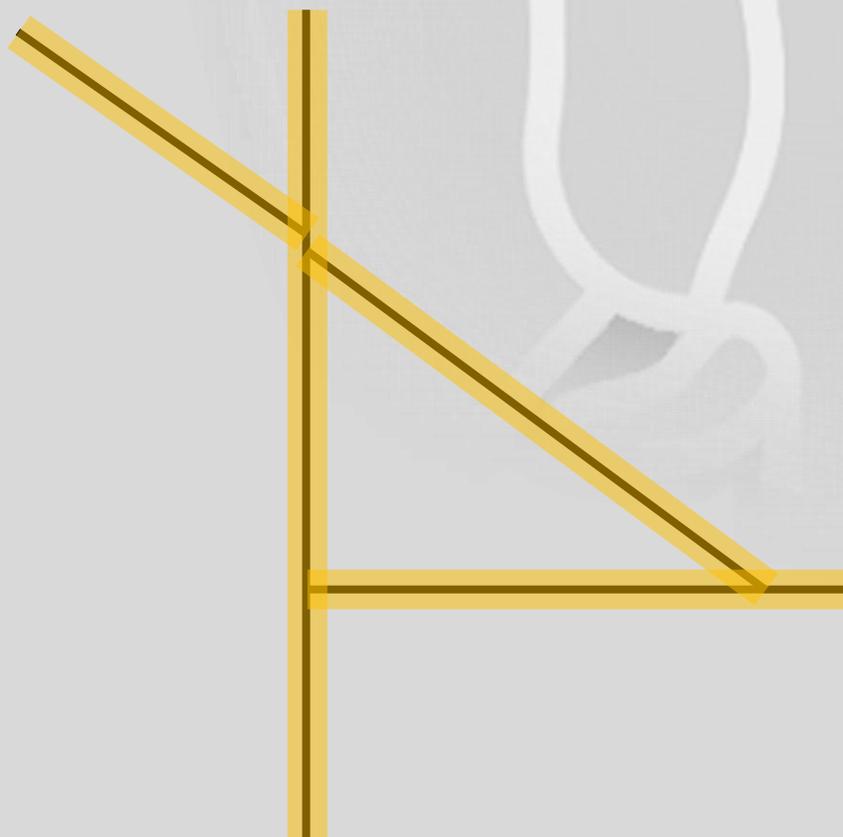


Linéaire 2



# Comparaison de 2 linéaires – calcul buffer

Linéaire 1

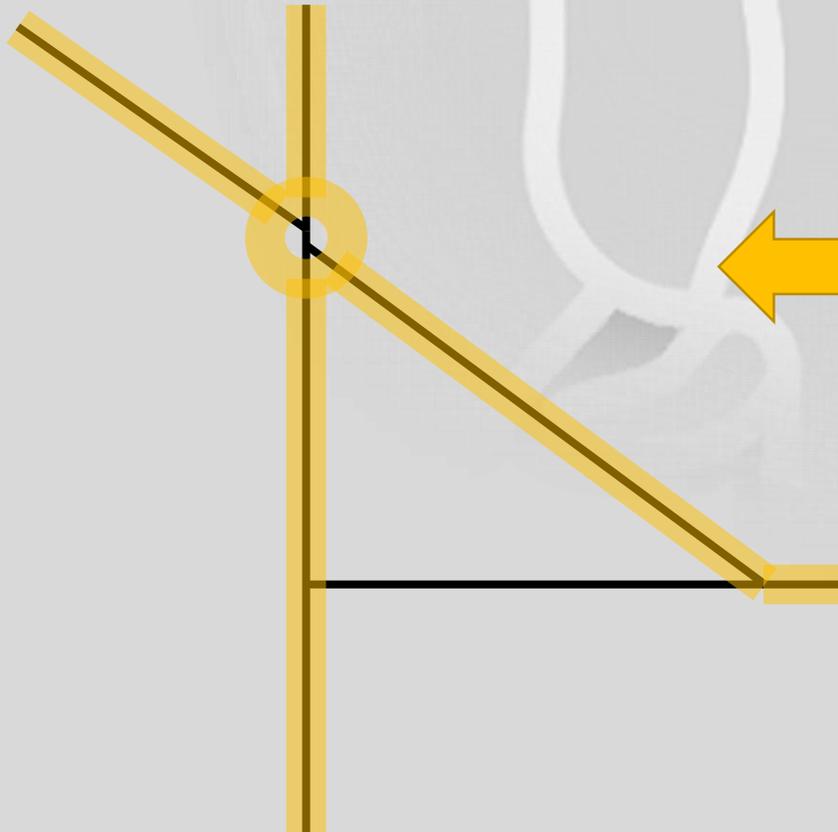


Linéaire 2



# Comparaison de 2 linéaires - superposition

Linéaire 1

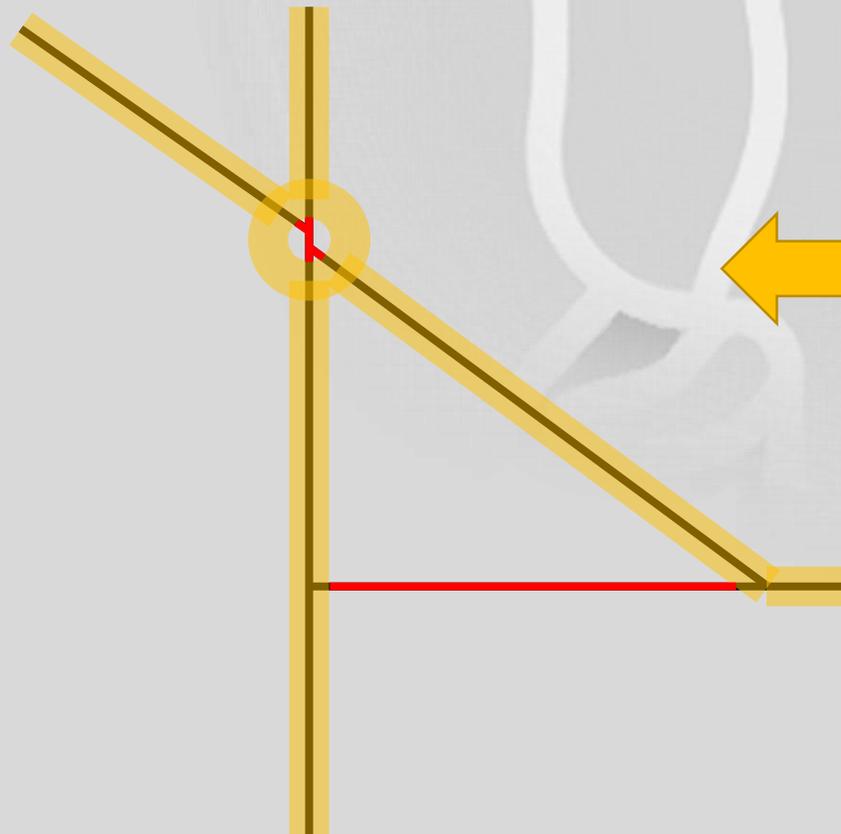


Linéaire 2

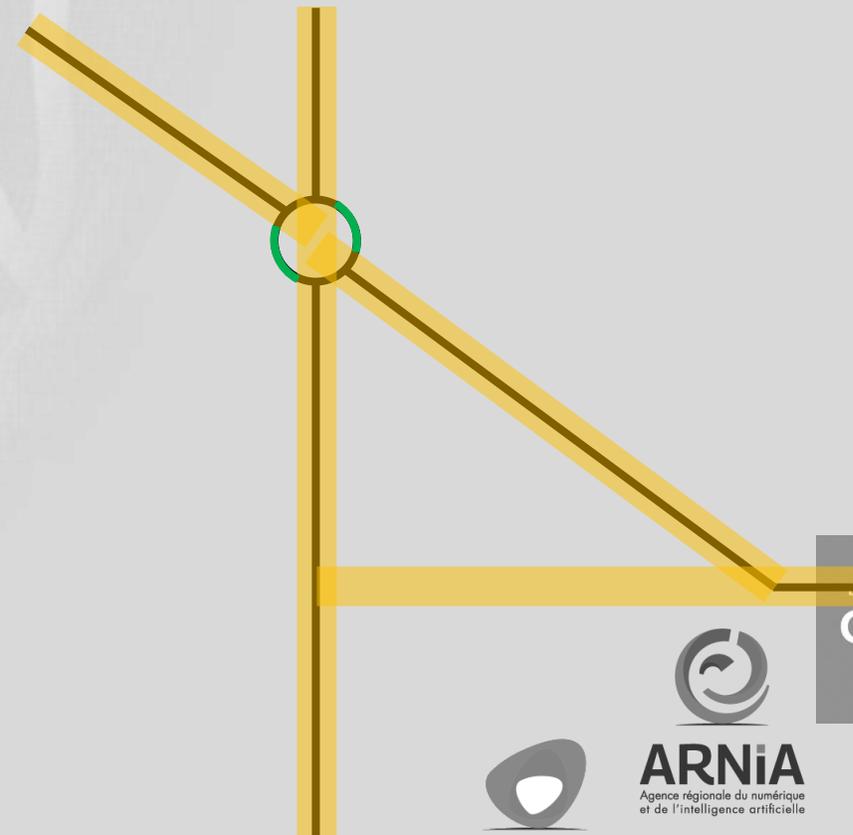


# Comparaison de 2 linéaires – analyse des écarts

Linéaire 1



Linéaire 2



# Comparaison de 2 linéaires - SOTM

Calcul d'un différentiel entre le référentiel routier du Gard et OSM - SOTM FR 2023

Replay : <https://peertube.openstreetmap.fr/w/7KFz9RF4XrFi4sY8ybrVWC>



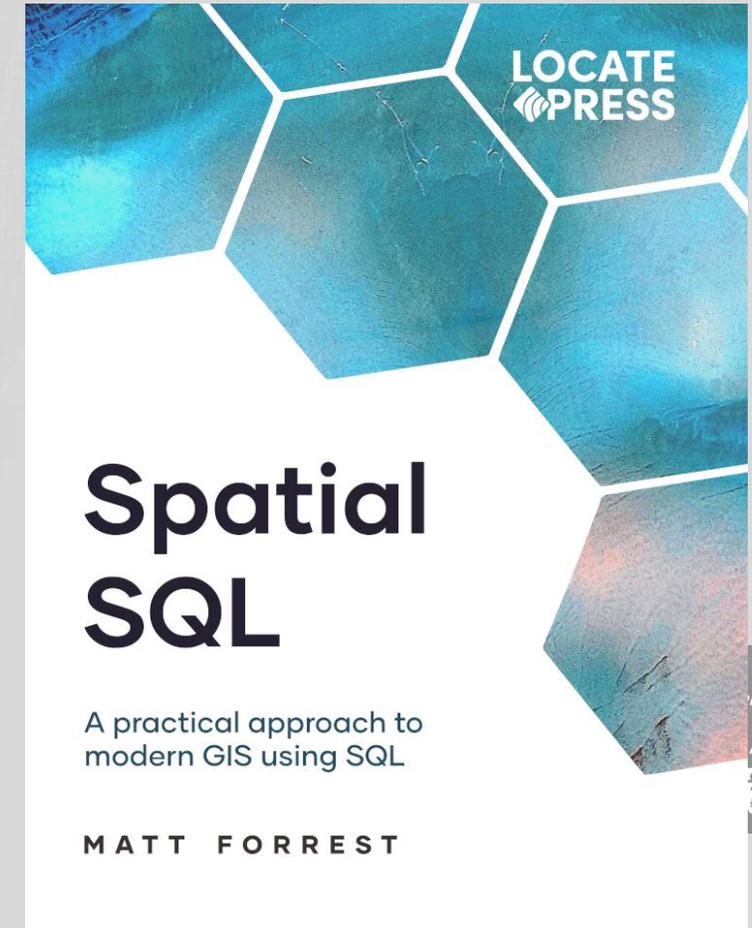
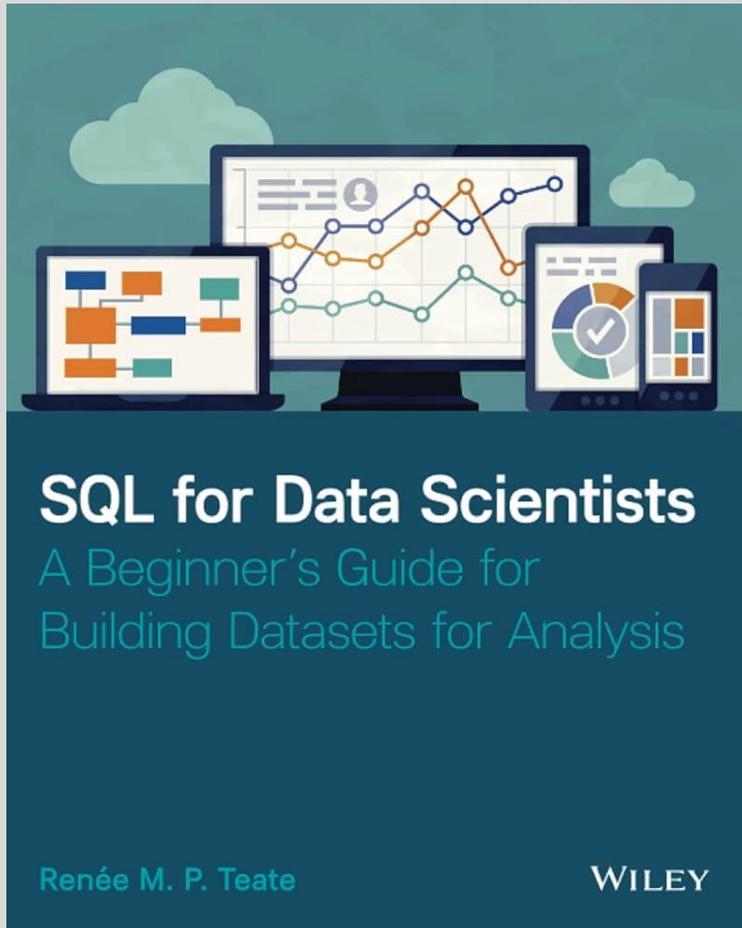
# #FollowFriday

- SQL
  - [@BecomingDataSci](#) (Renée Teate)
- PostgreSQL
  - [@crunchydata](#)
  - [@dalibo](#)
  - [@jpargudo](#) (Jean-Paul Argudo)
  - [@craigkerstiens](#) (Craig Kerstiens)
  - [@sqlliz](#) (Elizabeth Garrett Christensen)
  - [@l\\_avrot](#) (Lætitia Avrot)
- PostGIS
  - [@geotribu](#)
  - [@Data\\_Wax](#) (Florian Boret)
  - [@helenmakesmaps](#) (Helen McKenzie)
  - [@pwrapsey](#) (Paul Ramsey)
  - [@RustProofLabs](#) (Ryan Lambert)

# #JeudiLecture

- [How to SQL: a Guide for GIS Users](#) par [@helenmakesmaps](#)
- [5 maps you didn't know you could create with SQL](#) par [@helenmakesmaps](#)
- [Learn Postgres at the Playground](#) par [@craigkerstiens](#)
- [Intégrer les données OpenStreetMap dans son SIG pour s'engager dans un processus de contribution réciproque](#) par [@Data\\_Wax](#)
- [3DS: mesurer l'impact du transfert des Routes Nationales aux Départements](#) par [@Tetranos](#)

# #JeudiLecture



# Youtube

- [Replay PostGIS DAY 2022](#) sur la chaîne [Youtube Crunchy Data](#) (et bientôt 2023 ?)

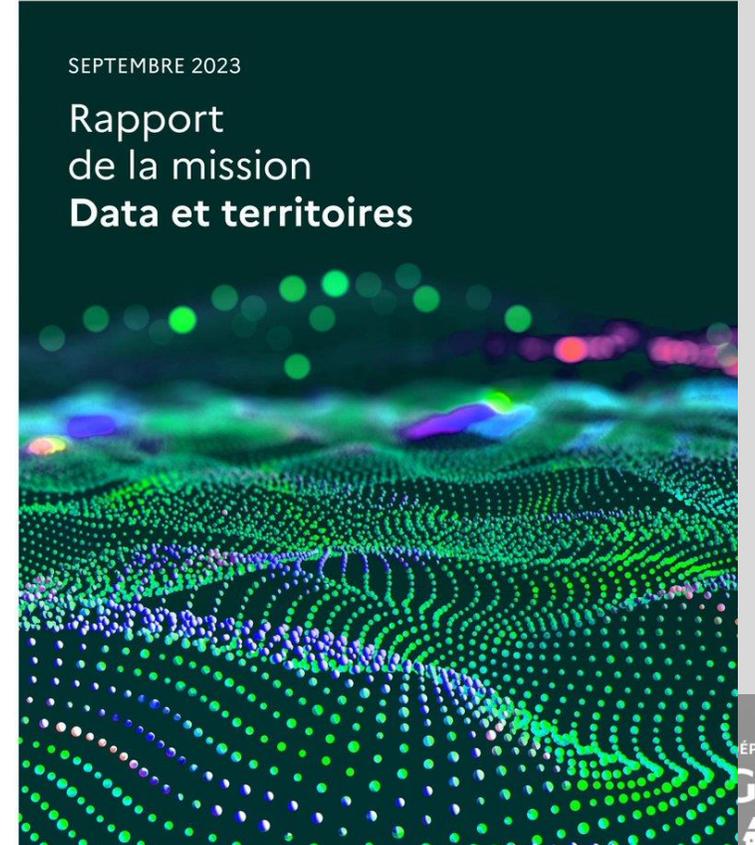
# DATA

- [@KahanDataSolutions](#)
  - [Modern vs Traditional Data Stacks](#)
  - [ETL vs ELT](#)
  - [dbt \(data-build-tool\)](#)

  
MINISTÈRE  
DE LA TRANSFORMATION  
ET DE LA FONCTION  
PUBLIQUES  
*Liberté  
Égalité  
Fraternité*

SEPTEMBRE 2023

## Rapport de la mission Data et territoires



CONSEIL  
DÉPARTEMENTAL  
GARD



**ARNIA**  
Agence régionale du numérique  
et de l'intelligence artificielle

## Fin de la session 3 et fin de la formation

C'était sympa !  
À refaire !



### Récap'

- Les possibilités d'analyse sont presque sans limite.
- Cela nécessite de connaître SQL + les fonctions propres à PostgreSQL (generate\_series) + les fonctions PostGIS (ST\_\*) + les autres extensions (unaccent, pg\_trgm).
- La connaissance exhaustive de l'ensemble est impossible, ce qui implique nécessairement une veille.
- Parce que « scriptées », les analyses sont facilement rejouables voire automatisables (approche as code).

# Des questions ?





pgQuaDoGéo © 2023 by Michaël GALIEN is licensed under CC BY 4.0.

To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

